UNIVERSITY OF CALIFORNIA, MERCED

The perceptual and cognitive roles of the motor system

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Cognitive & Information Sciences

by

Chelsea Leigh Gordon

Committee in charge:

Professor Ramesh Balasubramaniam, Ph.D., Chair Cognitive & Information Sciences
University of California, Merced

Professor Heather Bortfeld, Ph.D. Department of Psychological Sciences
University of California, Merced

Professor Michael J. Spivey, Ph.D. Cognitive & Information Sciences
University of California, Merced

2019

The work presented in this thesis was carried out in the Cognitive & Information Sciences (CIS) program at the University of California, Merced, U.S.A. Multiple chapters are the results of collaborations with UC faculty and alumni.

The dissertation of Chelsea L. Gordon is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Professor Ramesh Balasubramaniam, Ph.D.
Cognitive & Information Sciences, UC Merced

Professor Michael Spivey, Ph.D.
Cognitive & Information Sciences, UC Merced

Professor Heather Bortfeld, Ph.D., Chair
Cognitive & Information Sciences, UC Merced

University of California, Merced
2019

Contents

Lists of Tables and Figures

External Funding

- UC Music Experience Research Community Initiative (UC MERCI)

Internal Funding

- Graduate Dean's Dissertation Year Fellowship
- Eugene Cota-Robles Fellowship **(2 years)**

Collaborators/Co-authors

Marco Iacoboni	Dept of Psychiatry, UCLA
Timothy Shea	AI Labs, Accenture
Patrice Cobb	Dept of Quantitative Psychology, UC Merced
David C. Noelle	Dept of Cognitive and Information Sciences, UC Merced
Michael Spivey	Dept of Cognitive and Information Sciences, UC Merced

Personal Acknowledgements

I am indebted to my collaborators for igniting a deep passion for the research I pursued here. I am also eternally grateful to those friends, labmates, and family who supported me emotionally and encouraged me through rocky times.

Chelsea L. Gordon
PhD Candidate**,** chelseagordon.me
Cognitive and Information Sciences, 5200 Lake Road, University of California, Merced

EDUCATION

2013 - Cognitive and Information Sciences PhD Student, University of California, Merced
2008-2012 Psychology B.S., Michigan State University, Minor: Cognitive Science

PUBLICATIONS

Gordon, C. L., Shea, T. M., Noelle, D., Balasubramaniam, R. (accepted). Affordance compatibility effect for word learning in virtual reality.

Gordon, C. L.*, Hazam, P.*, Balasubramaniam, R. (2018). Recruitment of the motor system during music listening: An ALE meta-analysis of fMRI data. *PLoS ONE.* DOI: 10.1371/journal.pone.0207213 (in press: accepted 10/29/2018)

Gordon, C. L., Iacoboni, M., Balasubramaniam, R. (2018). Multimodal music perception engages motor prediction: a TMS study. *Frontiers in Neuroscience, 12,* 736.

Holbrook, C., **Gordon, C. L.**, Iacoboni, M. (2018). Continuous theta-burst stimulation of posterior medial frontal cortex to experimentally reduce ideological threat responses. J. Vis. Exp. (139), e58204, doi:10.3791/58204.

Gordon, C.L., Spivey, M.J., Balasubramaniam, R. (2017). Corticospinal excitability during the processing of handwritten and typed words. *Neuroscience letters*, *651*, 232-236.

Gordon, C.L., Balasubramaniam, R. (2017). Recruitment of the motor system in the perception of handwritten and typed characters. Proceedings of the 39th Annual Conference of the Cognitive Science Society. London, UK.

Bunce, J.P., Abney, D.H., **Gordon, C.L.,** Spivey, M.J., Scott, R.M. (2016). Using Motor Dynamics to Explore Real-time Competition in Cross-situational Word Learning: Evidence From Two Novel Paradigms. Proceedings of the 38th Annual Conference of the Cognitive Science Society. Philadelphia, PA.

Bunce, J., **Gordon, C.,** Abney, D., Fleming, M., Greenwood, M., Chiu, E., Spivey, M., Scott, R. (2015). Mouse tracking reveals knowledge of multiple competing referents during cross-situational word learning. Proceedings of the 40thAnnual Boston University Conference on Language Development. Boston, MA.

Gordon, C., Anderson, S., & Spivey, M. (2014). Font can change how we think about what we think. Proceedings of the 36th Annual Conference of the Cognitive Science Society. Quebec City, Canada

Sarinopoulos, I., Hesson, A. M., **Gordon, C.,** Lee, S. A., Wang, L., Dwamena, F., & Smith, R. C. (2013). Patient-centered interviewing is associated with decreased

responses to painful stimuli: An Initial fMRI study." Patient education and counseling, 90(2), 220-225

POSTER PRESENTATIONS

Gordon, C.L., Pabst, A., Balasubramaniam, R. (2019). A functional role for primary motor cortex in memory for handwritten and manipulable nouns. Annual conference of the Cognitive Neuroscience Society.

Gordon, C.L., Cobb, P., Balasubramaniam, R. (2018). Recruitment of the motor system during music listening: an ALE meta-analysis. Annual conference of the Cognitive Neuroscience Society.

Gordon, C.L., Balasubramaniam, R. (2017). Recruitment of the motor system in the perception of handwritten and typed characters. Proceedings of the 42th Annual Conference of the Cognitive Science Society. London, U.K.

Gordon, C.L., Balasubramaniam, R. (2017). Corticospinal excitability during the processing of handwritten and typed words. Annual conference of the Cognitive Neuroscience Society.

Gordon, C. L., Balasubramaniam, R., Iacoboni, M. (2016). Investigating modulation of cortico-spinal excitability in multi-modal music perception. 14th International Conference for Music Perception and Cognition, San Francisco, CA.

Gordon, C. L., Spivey, M.J., Balasubramaniam, R. (2016). Motor cortex excitability during processing of handwritten and typed non-action-related text. 38th Annual Conference of the Cognitive Science Society. Philadelphia, PA.

Bunce, J.P., Abney, D.H., Gordon, C.L., Spivey, M.J., Scott, R.M. (2016). Using Motor Dynamics to Explore Real-time Competition in Cross-situational Word Learning: Evidence From Two Novel Paradigms. 38th Annual Conference of the Cognitive Science Society. Philadelphia, PA.

Gordon, C., Anderson, S., & Spivey, M. (2014). Font can change how we think about what we think. Proceedings of the 36th Annual Conference of the Cognitive Science Society. Quebec City, Canada.

Bunce, J., Gordon, C., Abney, D., Fleming, M., Greenwood, M., Chiu, E., Spivey, M., Scott, R. (forthcoming). Mouse tracking reveals knowledge of multiple competing referents during cross-situational word learning. Proceedings of the 40thAnnual Boston University Conference on Language Development. Boston, MA.

INVITED TALKS

12/2018 The Inaugural UC Merced CogSci Symposium: "New Beginnings", UC Merced

08/2018 The Eleventh Embodied and Situated Language Processing Conference, Lancaster, UK

06/2016 Marco Iacoboni Systems Neuroscience lab, UCLA Ahmanson-Lovelace Brain Mapping Center

ADDITIONAL TRAINING

2016 UC Retreat Workshop on Research in Music Experience and Communication, Marconi, CA
2016 UC Music Experience Research Community Initiative (MERCI) Student Exchange. June 7th-14th, Ahmanson-Lovelace Brain Mapping Center, University of California, Los Angeles.
2015 UC MERCI Symposium/Workshop on Research on Music Experience and Communication, University of California, Los Angeles.
2015 Autumn School for the Sensorimotor Foundations of Social Cognition. Boltenhagen, Germany.
2015 MathWorks Seminar: Data Analysis and Visualization with MATLAB
2015 Transcranial Magnetic Stimulation (TMS) Methods and Practice Certification Course. Division of Biokinesiology and Physical Therapy, University of Southern California.
2015 APA Advanced Training Institute: Nonlinear Methods for Psychological Science. University of Cincinnati, Ohio.
2015 Generalized Linear Models Workshop. University of California, Merced. (by Bodo Winter)

AWARDS & FELLOWSHIPS

2019	Graduate Dean's Dissertation Year Fellowship, UC Merced
2017	GRAD-EXCEL Peer Mentorship Award, Graduate Division, UC Merced
2015, 2016	UC MERCI Rapid Research Grant
2014, 2015	CIS Travel Fellowship, UC Merced
2014, 2015	CIS Summer Fellowship, UC Merced
2013-2015	Eugene Cota-Robles Fellowship

TEACHING/PROFESSIONAL EXPERIENCE

Spring 2018	Graduate Division Statistics Tutor, University of California, Merced
Spring 2016, 2017	Teaching Assistant, Research Methods (Instructor: Michelle Greenwood, Rick Dale), University of California, Merced
Summer, 2017	Cognitive Neuroscience, University of California, Merced
Summer 2015	Graduate Student Researcher, Academic Placement Data and Analysis project Qualitative Measures advisor; funding by American Philosophical Association.
Fall 2015, 2016	Teaching Assistant, Perception and Action (Instructor: Ramesh Balasubramaniam), University of California, Merced.
12/2015	TMS workshop for BIO 113: Bioinstrumentation.

SERVICES AND VOLUNTEER WORK

2018 Lead organizer, Mental Simulation Workshop 2018, UC Merced
2016 Volunteer, 37th Annual Meeting of the Cognitive Science Society, Philadelphia, PA

2015	Public outreach, Mercy Hospital 5K Stroke Awareness Run
2015 –	Student Institutional Review Board Member
2014	Cognitive and Information Science Departmental Meeting Co-Organizer

AFFILIATIONS

Women in Cognitive Science
Society for Neuroscience
Society for Music Perception & Cognition
Cognitive Science Society

Abstract

The motor system in the brain is crucial in allowing us to successfully move around in our environment, interact with people and objects, and execute finely controlled motor commands. While most of the early neuroscience research on these regions tends to focus on these "main" functions, over the last few decades evidence has been surfacing that points to a more broadly integrated role for the motor system. Many recent findings suggest that it is also of importance in many other aspects of human cognition, from language and thought to social cognition and, as I discuss in depth in the following sections, many perceptual processes. In the following chapters, I outline and compare existing prediction-based and simulation-based theories for motor system involvement in perception. I also describe experiments I completed investigating motor system involvement in written language perception, music perception, and action observation. Furthermore, I discuss how these processes relate to conceptual learning and recall. In summary, a vast literature points to the motor system proper not being a neural network that is only good for controlling and planning our actions. As we develop the vocabulary of the field to use terms like "action-perception loops" and discuss these processes as less separable than previously considered, perhaps we should also reconsider the term "motor system" to reflect its diverse roles in sensorimotor prediction.

Chapter 1
Motor system involvement in perception

1. Introduction/what is the motor system?

What neuroscientists call the motor system is comprised of many different brain regions that come together to produce action. The motor system consists of cortical regions in addition to subcortical structures (i.e., the cerebellum and basal ganglia) and the brainstem and spinal cord. The neurons primarily responsible for causing muscles in the effectors to move are the alpha motor neurons located in the spinal cord. These control simple reflexive movements and receive a lot of feedback from cortical structures issuing motor commands. The pathway from cortical motor system areas to those located in the spinal cord is called the corticospinal tract. The fibers in this tract originate in primary motor cortex (M1), where simple muscle-specific motor commands are issued, as well as multiple secondary motor areas, including premotor cortex (PMC) and supplementary motor areas (SMA).

Each of these motor regions contain what is called a *somatotopic map* of the body. In M1, this map corresponds to particular muscles of the body, with more cortical surface allocated to those muscles used often and for precise movement. The secondary motor areas contain somatotopic maps that correspond not to single muscles, but to behavior-related muscle groups, for instance, the muscle group corresponding to a grasping action. The premotor cortex is further divided into the ventral (vPMC) and dorsal (dPMC) subregions, which exhibit differences in the information they specialize in processing. vPMC is strongly implicated in speech production and grasping behaviors. Recent evidence points to a role for dPMC in switching motor plans online. The primary role for SMA is that of motor sequence learning, or other self-guided movements, while the regions of PMC specialize in sensory-guided movement.

Other less central cortical areas in the motor system include the somatosensory cortex (S1) and posterior parietal cortex (PPC) which also contain somatotopic maps of the body. S1, receiving extensive sensory information from the thalamus, is important for interfacing sensory information with motor plans. PPC, whose primary role is processing spatial information and sensorimotor transformations, is crucial for moving in space and the environment.

The cerebellum is involved in many motor-related functions, perhaps not all of which are yet discovered. Parts of the cerebellum contribute to balance, fine motor control, and eye and head movements. Other regions are implicated in timing mechanisms as well as online error-correction processes in movement. The other subcortical motor region, the basal ganglia, has implications for motor learning and for initiating and inhibiting motor commands.

The motor system in the brain is crucial in allowing us to successfully move around in our environment, interact with people and objects, and execute finely controlled motor commands. While most of the early neuroscience research on these regions tends to focus on these “main” functions, over the last few decades evidence has been surfacing that points to a more broadly integrated role for the motor system. Many recent findings suggest that it is also of importance in many other aspects of human cognition, from

language and thought to social cognition and, as I discuss in depth in the following sections, many perceptual processes.

2. Motor system involvement in perception

As mentioned above, many studies have shown that the motor system is active not only during action execution, but also during many instances of perception. Here I will discuss many of the perceptual tasks found to make use of the motor system. In doing so, I will also try to draw on commonalities and differences among the perceptual tasks that are studied. In section 3, I will review theories of motor involvement in perceptual processes and attempt to integrate the various theories and the relatively disparate literature described in this section.

2.1 Language perception

Many findings in the speech perception literature suggest a tight coupling between the motor process involved in speech production and that of speech perception. The motor theory of speech perception (Liberman & Mattingly, 1985) states that the representations of speech sounds involve a mapping of the particular sound to the motor articulations required to produce it. In other words, speech perception is accomplished via an understanding of speech production. This theory has been very influential in the field, although more recent evidence points to these mappings being a learned phenomenon instead of an innate mapping, as was proposed in the original work.

2.1.1 Developmental findings

Infant speech learning involves a very tight coupling between auditory and motor regions. The babbling stage of speech learning (between 5-10 months of age) is thought to promote the development of a link between articulations and their corresponding speech sounds, providing a basis for developing a phonetic inventory (Vihman, 1991). Bi-directional links between speech audition and production are seen in infants developing speech skills. For instance, it has been shown that deaf infants show impairments in babbling, and have subsequent impairments on intelligible speech, indicating the importance of auditory information for successful development of speech production (Oller & Eilers, 1988; Wallace, Menn, & Yoshinaga-Itano, 1998). Effects of speech production on speech perception abilities in infants have also been shown. Bruderer et al. (2015) had 6-month-old non-speaking infants perform a discrimination task on nonnative speech sound distinction. The authors used teething toys in the infants' mouths to constrain the movement and positioning of the tongue so that particular speech sounds would not be possible to create. They found that infants performed significantly better (measured by looking-time) on discrimination of speech sounds that were not made impossible by the teething toys, suggesting that even in pre-speech infants, the sensorimotor information from the articulators is crucial in speech perception.

2.1.2 Neuroscience findings

Many neuroimaging studies provide evidence for a link between speech production and speech perception. In an fMRI experiment, Wilson et al. (2004) found that adults listening to speech show activation in bilateral vPMC during speech listening.

Skipper, Nusbaum, and Small (2005) extended this fMRI experiment to be audiovisual by having subjects hear and watch talking faces. Increased activation was found in STS, pars opercularis, PMC, M1, S1, and cerebellum. Fadiga et al. (2002) found that while subjects are passively listening to the speech of another, there is facilitated corticospinal excitability in the tongue motor region of the brain. They used single-pulse transcranial magnetic stimulation (TMS) to probe excitability while Italian-speaking subjects listened to italian words and pseudo-words that contained either the "ff" or "rr" phoneme. The "rr" sound recruits the tongue muscle much more than the "ff" sound. The authors found that corticospinal excitability measured as motor-evoked potentials (MEPs) from the tongue muscle was higher while subjects heard the "rr" words than the "ff" words. In other words, the words that require more tongue movement to pronounce also activate those muscles more during the perceptual experience of listening. Interestingly, they also found that words induced a further increase in excitability over nonwords, suggesting an effect of whether the participant has performed the exact speech sounds before.

Watkins, Strafella, and Paus (2003) replicated these findings, but with MEPs recorded from the lips. They also extended this work to show that the visual observation of speech acts without auditory accompaniment facilitates corticospinal excitability. A control condition was also included, where stimulation/recording was of the hand area, to see whether there was a non-specific increase in excitability over regions of the motor cortex, not localized to the area representing the appropriate effector. No excitability (measured by increase in MEPs) was observed in the hand area during this control experiment, confirming that it was excitability specific to the appropriate effector region (the mouth muscles in this case).

Taken together, these experiments provide evidence that there is concurrent activation of effector-specific motor regions during perception of speech acts, both visual and auditory. What is not clear from these studies, however, is whether the role of motor activity in speech perception is causal, rather than epiphenomenal. While this has yet to be shown in primary motor cortex, research does suggest a causal role in premotor cortex. Meister et al. (2007) performed repetitive transcranial magnetic stimulation (rTMS) over the premotor cortex while participants performed a phoneme discrimination task. The task was to identify, amongst a steady background level of white noise, whether a syllable was pa, ta, or ka. The authors found that rTMS disrupted subjects' performance on this task, but not on an equally difficult color discrimination task. This provides evidence that the premotor cortex has a causal role in speech perception. In Chapter 2, I will describe a study showing that written language is also embodied in the motor system.

2.1.3 Findings from computational models

Westermann and Miranda (2004) designed a computational model that learns the sensorimotor coupling of motor parameters and sensory consequences that occur during babbling. The model consists of a motor command map and an auditory stimuli map, and can do one of two things: (1) it can babble, by generating a motor command and listening to the sound produced, or (2) it can listen to external sounds. With experience, the model develops connections between the motor and auditory maps, resulting in the fine tuning of the representations. The first experiment using this model explored the learning of the perception-production coupling and its effects on perception and production during

babbling. In this experiment, the model babbled by selecting a motor command and its associated auditory consequence. They found that after the babbling training, both the auditory and motor maps had developed clusters of preferred responses for linear regions of the motor sound mapping. In other words, these clusters are in regions that are easily reproducible and easily predicted. This is where small changes in the motor parameters lead to small changes in the produced sound. In a second experiment, the authors varied the language environment to include two different language sounds, German and French vowels. The model underwent training as it had in the first experiment and again developed clusters of prototypical sounds. These prototype clusters were very close to representing the actual French and German vowels. This model shows how a system with interacting perceptual and. Motor maps could give rise to an underlying substrate that encodes both action and perception.

The above literature suggests that various motor regions are active during speech perception, including vPMC, M1, S1 and cerebellum. Further, a causal role is suggested for vPMC in speech perception, or at least for discriminating phonemes. I have also reviewed developmental research that strongly indicates a central role of the motor system in the development of speech perception and production abilities. Next I will turn to the observed involvement of the motor system during music perception.

2.2 Music perception

Numerous studies also report motor activation during music listening tasks. Similarly to speech acquisition, learning to play music involves tight sensorimotor coupling of perception and production of music. Unlike speech, however, it is not necessary to acquire musical abilities in order to listen to music. This gives us a unique opportunity to examine differences between expert musicians, who have acquired this strong sensorimotor coupling in a musical domain, and non-musicians who have not.

Haueisen and Knosche (2001), in a magnetoencephalography (MEG) experiment, looked at brain activation in pianists and non-pianists while listening to piano pieces. They found that musicians exhibited significantly greater activity in primary motor cortex (M1) during music listening than non-pianists. In particular, they were able to find a dissociation between regions of M1 that control the thumb and little finger and found increased activation in the respective region depending on which of these digits would be used to play the perceived piano note in the piece. This suggests that the motor program corresponding to the perceived sound is active during the perception of the sound. Increased activation was not seen in other motor regions, such as SMA or PMC. The authors speculate that this is because motor planning is not needed in the highly automated motor programs that exist in the skilled musicians.

In a similar fMRI experiment, Baumann et al. (2007) searched for regions of the brain that were active in both passive listening and music playing, in expert pianists and non-musicians. The authors found activation during passive listening in dorsal premotor cortex (dPMC), the planum temporal (PT), and posterior parietal cortex (PPC), but not consistently in primary motor cortex. The authors conclude that these regions are a part of the substrate underlying integration and transformation across modalities. Motor activation (in dPMC) is suggested to underlie audio-motor transformation circuits, converting modality specific information into a more general form of information. Other

researchers have suggested similar roles of this neural circuitry (Doupe and Kuhl, 1999; Hickok and Poeppel, 2000). Musicians in the experiment conducted by Baumann et al showed greater activation than non-musicians during music listening in dorsal premotor cortex (dPMC) and pre-supplementary motor areas (preSMA), suggesting a role for these regions specific to acquired sensorimotor experience in music listening. Baumann and colleagues speculate that the reason primary motor cortex (M1) activation was not observed consistently in their listening tasks, while other studies have found M1 activation in music listening, is due to poor spatial resolution in the EEG and MEG experiments reporting this activation. They suggest that the activation detected in the other studies is likely coming from premotor cortex, but misinterpreted as coming from M1.

Many more experiments show similar results. Meister et al (2004) scanned participants in an fMRI experiment where they either played music on a silent piano or imagined playing on a piano. Increased activity in SMA and PMC was observed during both conditions. Activation in M1 was only observed in the silent playing condition, however. Bangert et al (2006) executed a similar experiment, using both musicians and non-musicians, and found a greater activation in SMA and PMC in the musicians during passive listening to piano melodies, replicating the results obtained by Baumann and colleagues.

An interesting instance of music perception in particular is beat perception. Beat perception is the detection of the pulse, or beat, underlying a musical piece. Motor activity is also observed during beat perception in multiple studies. Unlike the passive observation of a musical note created by a particular motor command, beat perception does not involve a mapping between a motor gesture and its corresponding sound. Rather, beat perception involves detection of a temporal component underlying the musical piece. This allows observation of specific motor contributions to rhythmic auditory perception. Beat perception has been shown to multiple regions of the motor system. Bengtsson and colleagues (2009) ran an fMRI experiment where participants passively listened to rhythmic or random sequences. Increased activity during the rhythmic sequences compared to the random ones was found in dPMC, SMA, preSMA, and cerebellum. In a similar study, Chen, Penhune and Zatorre (2008) looked at fMRI activation in participants who were listening to a musical rhythm in anticipation of having to tap that rhythm shortly after, and in participants who were unaware that they would later tap along with the rhythm. The same motor regions (SMA, PMC, and cerebellum) were active in both conditions, suggesting an automaticity of the perception-motor link and that motor planning is not necessary for motor activity corresponding to rhythm perception.

Grahn and Brett (2007) also found fMRI activation in SMA during beat perception in a music listening task. Participants in this experiment listened to either simple rhythms with a predictable beat or complex, not easily predictable rhythms. In particular, fMRI activation in pre-SMA/SMA was higher in the simple rhythms, where identifying the underlying beat is easier. Relatedly, the amount of groove (defined as an underlying sense of wanting to move to a piece of music) in a musical piece has been found to modulate corticospinal excitability of the motor system. While recording MEPs from the first dorsal interosseus (FDI) muscle in the hand during a TMS experiment,

Stupacher et al. (2013) observed an increase in MEP amplitude for high groove music over low groove music in musicians, and a decrease in MEP amplitude for high groove music in non-musicians. The authors conclude that high-groove music engages the motor system, and that musical training influences this process. They suggest that motor suppression occurring in non-musicians because of the instructions not to move during the experiment. In Chapter 3, I will describe a meta-analysis we did to find which areas the motor system are involved during passive music perception.

2.3 Action perception

The most studied perceptual domain with motor involvement is that of the perception of others' actions. Many findings in this area suggest that the motor system plays a crucial role in perceiving the actions of others. Many of these experiments show that the more the observer is capable of performing the observed action, the higher the activation in motor regions of the brain, suggesting that there is a mapping of the observed action onto one's own motor repertoire.

Fadiga et al. (1995) measured corticospinal excitability of human subjects passively observing either a dimming light, an experimenter interacting and grasping objects, an experimenter tracing figures in the air, or the objects alone. What they found was a greater increase in MEPs (area under the curve) during the conditions where the experimenter was moving his arms in comparison to the other conditions, indicating larger motor involvement during these trials.

Facilitation of corticospinal excitability of the hand motor area is also found when subjects listen to hand-related sounds. Aziz-Zadeh et al. (2004) used TMS to probe corticospinal excitability while subjects heard typing sounds, the sound of paper tearing, walking sounds or the sound of thunder. Significantly larger MEPs were observed during the sounds related to hand actions (i.e., typing and paper tearing), providing evidence that auditory action perception also activates the motor system. Importantly, this increase was only seen when stimulation occurred in the left hemisphere and recorded from the right hand. The authors interpret this as action observation being an important precursor to language, which takes place mostly in the left hemisphere.

The activation observed in PMC is found to be organized somatotopically, similar to the somatotopy underlying planned action execution. Buccino et al. (2001) found increased fMRI activation in the respective areas of PMC during the observation of actions performed with the hand, mouth, or foot effectors. This suggests that not only is the motor system active during action observation, but it is active in very similar ways to that during actual action execution. In another fMRI experiment, Calvo-Merino et al. (2005) asked participants in the scanner to passively observe videos of expert performers of ballet enacting ballet moves or of expert capoeira performers enacting capoeira moves. Participants were either expert ballet dancers, capoeira dancers, or neither. For the ballet dancer subjects, greater brain activation was observed in PMC during the observation of classical ballet moves than during the capoeira videos; the opposite pattern was found for the capoeira dancers. As for the non-dancers, the two conditions showed the same increase of PMC activation.

Common aspects underlying both observed and executed action are visual properties of the action. In other words, there are similar underlying visual components to seeing one's own hand reach for a glass, and seeing another individual's hand reach for a glass. Calvo-Merino and colleagues (2006) explored whether this visual component could explain the overlapping action in motor regions or whether the motor component underlies the overlap. They also had expert ballet dancers in the fMRI scanner, but they separated the amount of visual experience a dancer had with a dance move and the amount of motor experience they had with those same moves. Since ballet dancers often practice and perform with a counterpart of the opposite gender, the authors had male and female ballet dancers observing male and female dance moves. While either gender only experiences their gender-typical dance moves, they have extensive visual experience of the moves enacted by their opposite gender counterparts. The study found a stronger increase in premotor BOLD activity in male subjects observing male dance moves and in female subjects observing female dance moves, suggesting that it is in fact the motor experience that leads to greater activation of the cortical motor regions. The authors also found the same differential activation pattern in the cerebellum, providing evidence that the cerebellum has a role in the action-observation network. In Chapter 4, I will discuss an experiment looking at simulation and prediction processes during multimodal music perception.

2.3.1 Cognitive engineering via principles of action-observation

The well-known research on the shared substrate for action execution and observation has even been harnessed in therapeutic contexts. Ertelt et al. (2007) created something they called *action observation therapy* for patients suffering from recent strokes experiencing difficulty regaining control over the movement of their limbs. In addition to the physical training normally implemented, some of the patients in this study received a training where they observed actions made by others. Compared to a control group who did not receive this new additional training, the experimental group had significantly faster recovery of motor function. Post-study fMRI scans revealed that this group also yielded a significant increase in motor activity, both in vPMC and SMA.

The action-observation literature has also influenced some of the approaches to creating functioning robotics. Baraglia et al (2015) built a computational model using a recurrent neural network that integrates visual and motor information. The system's visual attention used a calculated prediction error between predicted and actual sensory outcome values. Using a simulated humanoid robot, they found that the production of self-generated actions biased perception of observed actions performed by others. This highlights the influential role of visuomotor experience on action perception.

2.4 Other action-related perceptual processes

McGettigan et al. (2013) conducted an fMRI experiment where subjects listened to authentic amusement laughter and to controlled, voluntary laughter. Activation in pre-SMA was observed for both types of laughter. Furthermore, an individual differences analysis showed that pre-SMA activity in a subject correlated with the subject's accuracy in classifying the laugh as either authentic or voluntary. This suggests that motor activation during laughter perception is potentially beneficial for social cognition.

In an fMRI study with expert and novice athletes, Woods et al. (2014) had subjects listen passively to sports sounds (tennis and basketball) and non-sports sounds. Non sports sounds consisted of common sounds such as a flushing toilet or crumpling paper. The sports sounds were of a basketball bouncing at varying intensities and of a tennis ball being hit by a racquet, and experts were only experienced with one of the two sports. They found that all athletes showed activation in bilateral SMA, left precentral gyrus, and bilateral postcentral gyrus while listening to sports sounds. Expert athletes showed the highest activation in inferior frontal gyrus (IFG), which has been linked to motor planning processes, during the perception of the sounds associated with the sport they were familiar with. The authors suggest this is due to the athletes' sensorimotor system being attuned to highly relevant sounds and tuning out irrelevant sounds. In novices, greater activation in medial frontal gyrus (MFG), associated with motor processing and imagery, was found for the sounds of the sport they were *less* familiar with. This might involve a strategy of using the sensorimotor system to identify the target of the sounds.

2.5 Non-biological event perception

Wilson and Knoblich (2005) argue that motor activation during the perception of actions performed by conspecifics is largely due to its role in perceptual prediction. However, one emerging line of research coming from Ricarda Schubotz and colleages looks at activation of motor regions during the perception of non-biological event stimuli. The authors suggest (Schubotz, 2007) that the perceptual prediction in motor cortex is not limited to the movement of humans or even biological animals, but any perceptual prediction. We are unable to imitate inanimate events, such as the rolling of waves or a flashing light, in any practical sense such as how we might imitate human actions. We can, however, use the spatiotemporal information present in an event to anticipate how it might unfold, in a similar way to how we might anticipate human action. For instance, we can anticipate when a rolling wave will reach its peak and begin to fall.

Schubotz and von Cramon (2004) designed an fMRI experiment where subjects predicted the expected outcome of observed actions, motor imagery, or geometrical figure sequences. Increased activation in motor regions was found for all of these conditions when compared to a basic sequential target detection task. Regions that were active primarily during biological stimuli and regions active primarily during nonbiological stimuli were also discovered. During the biological stimuli conditions, activation was seen in inferior frontal gyri (IFG), the extrastriate body area, posterior temporal sulci, and somatosensory cortices. The active areas in the nonbiological stimuli conditions included pre-SMA, middle frontal gyri (MFG), intraparietal sulci, and the caudate nucleus of the basal ganglia.

In a meta-analysis of fMRI experiments, Schubotz and von Cramon (2002) show that several types of serial prediction tasks (object, spatial, and rhythm/pitch) activate premotor areas in a somatotopic way. In particular, spatial tasks activate parts of premotor cortex that are normally active in the observation and execution of foot and arm actions, object tasks activate regions associated with hand and wrist action observation/execution, and pitch/rhythm tasks activate regions that control mouth and finger movements.

3. Theories for motor system involvement in perception

In this discussion, I will reference a few of the current proposals for the mechanism behind motor system involvement in perceptual tasks. I will do this by first highlighting the commonalities between all of these different types of perceptual processes, and thus, what could potentially be a common role for the motor system in each. Second, I will discuss the differences between these perceptual processes and the potential for motor system activity in each of these modalities to be a result of different regional networks recruited for particular perceptual domains. The existing theories of motor involvement in perception are disparate and incomplete at this point. Including all of these theories in one place will enable a comparison between them and the potential to combine together the aspects that fit into a cohesive whole.

3.1 Simulation theories

3.1.1 Mirror Neurons

The rapid burst of research investigating motor system activation during action observation is largely due to the field turning discovery of mirror neurons (MNs), which are neurons found in F5 of macaque monkeys (the monkey homologue of human vPMC) that fire during both the observation and the execution of actions (Rizzolatti et al., 1988; di Pellegrino et al., 1992; Gallese et al., 1996). In particular, this overlapping activation for observation and execution is hypothesized to be the underpinning for action understanding in humans. Rizzolatti and colleagues recorded from 532 neurons in area F5 and found that 92 of them (~18%) classified as mirror neurons. Out of these neurons, 30% were what they called "strictly congruent", meaning they responded only to the same action (i.e., grasping, reaching, etc.) with the same execution of that action. These neurons were further classified into types based on the action they responded to, which included grasping neurons, placing neurons, manipulating neurons, hand interaction neurons, and holding neurons. The rest of the mirror neurons were considered "broadly congruent", as they responded to more than one action and execution of those actions. A larger subset of all of the neurons recorded from were neurons that responded to objects, called canonical neurons. Rizzolatti and Fadiga (1998) proposed that the commonality between motor neurons and canonical neurons is that both of these types "generate an internal copy of a potential hand action." In other words, canonical neurons provide information about the graspable property of an object, while mirror neurons provide information on actions performed by another person.

While mirror neurons have not been reliably directly observed in humans (though, see Mukamel et al. (2010) potential findings of MNs in SMA and hippocampus), research suggests that they may exist in many regions in the human cortex. Experiments using fMRI and PET have shown regions in the human premotor and parietal cortices that respond both to observed and executed actions, leading researchers to suggest an analogous system in humans (Aziz-Zadeh et al., 2006; Buccino et al., 2004; Grezes et al. 2003). There is large debate on the proposed functionality of the mirror neuron system

and its role in action understanding (Hickok, 2009) and alternative interpretations of MN function abound (Csibra, 2008; Cook et al., 2014; Brinckner, 2010).

3.1.2 Jeannerod's simulation theory

Jeannerod (2001) proposes a purely simulation-based role for the motor system, suggesting that all covert actions are the same, neurologically, as overt actions that are not executed. He refers to mental states termed *S-states*, that exist during simulations of executed actions. He defines a number of S-states, differing by degrees of awareness of the simulation. For instance, the S-state for "imagined action" is conscious, whereas the S-state for the passive observation of graspable objects is non-conscious. The observed underlying neural activity is overlapping for each of these types of S-states, but differs in observable ways as well. Jeannerod reports studies showing activation corresponding to S-states in primary motor cortex, the corticospinal pathway, the basal ganglia, cerebellum, and premotor cortex.

Jeannerod describes two potential explanations for why covert actions would not lead to real actions. One is that motor activation during these S-states is subliminal and insufficient to cause the firing of actual motor neurons. The other potential explanation is that actual motor output is blocked by counteracting inhibitory mechanisms generated at the time of the motor command. The theory is mostly underspecified from a neuroscientific perspective, and falls short of offering an explanation for all of the cases of motor involvement in perception mentioned above.

3.2 Emulation and prediction theories

Some recent accounts of cognitive, perceptual, and motor activity in the brain work under the assumption that some kind of predictive process underlies such activities. The strongest proponent of this is Clark (2016), who suggests that all processing in the brain is predictive. The emulator framework originally put forward by Grush fits into this realm of theories, as does the more recent extension of this proposed by Wilson and Knoblich (2005), and Schubotz's even more encompassing suggestion that these predictive models might extend to non-biological events.

3.2.1 Grush's emulator theory

Grush (2004) put forth a theory called the "emulation theory of representation", where he develops a framework based on emulators in the brain used for representation in motor acts, in addition to motor imagery and other cognitive tasks. Grush does not advocate for the role of these emulators in the motor system, in particular, but the possibility is not ruled out. An emulator is a kind of forward model of a system that implements the input-output functions of the system and implements a Kalman filter to weight sensory/motor components. The forward models (referred to here as emulators) work similarly to those in some theories of motor control (Wolpert et al, 1995). A "controller" (potentially the motor system) issues a motor command, the "plant" (the effectors) receives this motor command and the action is executed (referred to as the input-output function of the plant). A copy of the sent motor command, called an efference copy, is sent to an *emulator*, which implements the same input-output function as the plant. This serves as a sort of prediction of the sensory information resulting from

the motor command. This is then sent from the emulator back to the controller, on a faster timescale than sensory feedback from the peripherals of the body (or plant). This information is then used to make online corrections or adjustments to the ongoing motor command, as well as to compare actual sensory feedback with that predicted by the emulator.

Grush goes on to describe how the emulation theory can account for processing in many areas outside of motor control. For mental imagery, for example, he suggests that the gain on the filter is set to have no influence of real sensory information and the actual motor command must be suppressed. With these factors removed, mental imagery can be carried out via emulators without having any overt movement of the body. He extends this theory to encompass perceptual processes as well. Perception relies on what he calls *environment emulators*, which track the current environment and objects in the environment, anticipating changes in the environment that could arise either based on the dynamics of the objects in the environment or based on movements produced by the observer.

Wilson and Knoblich (2005) expand on Grush's concept of an emulator extended to visual perception of actions by conspecifics via prediction processes. They hypothesize that "the perceptual system develops a structural internal model of the system to be predicted, a model that is isomorphic on a part-by-part basis to that external system, and contains information about the mechanics of its movement properties". These emulators contain representations of the principles and regularities of the system, allowing fast prediction of the system behavior. Because we have an internalized model of our own system/bodies and its principles and regularities, we can use this model for perceiving the actions performed by others who share similar bodies with similar principles and regularities.

3.2.2. Schubotz's theory of action perception of non-human events (the HAPEM framework)

Schubotz (2007) extends the idea of emulators to encompass inanimate event perception in addition to human action prediction. Although non-human events, such as a wave rolling or insects flying, are unable to be modeled using an emulator of our bodies they are, to a reasonable extent, predictable as they perceptually unfold. In this theory, we use our sensorimotor system and the models underlying simulation processes in order to make online predictions about observable events as they unfold.

How does this predictive process work? Schubotz suggests the following: when we repeatedly hear a melody, the lateral premotor cortex builds up sensorimotor representations using input from association areas of the cortex. These sensorimotor representations are only audiomotor, lacking the proprioceptive-motor representations that are involved in an actual performed movement. The lateral premotor cortex eventually establishes an internal model of this melody which can be used for perceptual prediction. This internal model is similar to that involved in motor control, but with the components for movement and sensory feedback removed.

Schubotz proposes what she calls the HAPEM (Habitual Pragmatic Event Map) framework, which states that "the prediction of an event that is structured with regard to a property P engages the area of the lateral premotor cortex that is best adapted to specify

its motor output in terms of property P". What this means is that the perception of events with different properties recruits particular somatotopic regions of vPMC, selected based on similarity to the underlying properties of that area of vPMC. For instance, the regions of vPMC that correspond to executing and observing mouth movements would be recruited for the perception of rhythmic events, due to the underlying rhythmic nature of the vocal system. Indeed, in their meta-analysis, this is exactly what they see.

3.3 Neural Reuse Theory

In his book "After Phrenology", Anderson puts forth an argument in favor of the theory of neural reuse underlying brain function. Under this framework, neural areas are reused for tasks that involve similar computational functions. For example, the region of the brain associated with finger gnosis (or the awareness of the position of the fingers and distance between them) also shows increased activation during numerical processing tasks. Anderson suggests that the neural structures originally evolved for finger awareness were then repurposed for the later acquired function of numerical representation. Importantly, the theory does not assume that any new function or task is using the *function* associated with the neural structure (in other words, it is not the case that numerical processing is *using* finger gnosis as a medium). Instead, it assumes that there are particular computational properties that regions of the brain are more accommodated for, and these computational properties are exploited for the relevant tasks or cognitive domains.

Under this framework, what we look for as an explanation for motor recruitment in perception and high-level cognition might be much simpler and basic than that of theories like simulation. Perhaps the structure of the motor system is adaptive for computational processes underlying many of the functions in cognitive and perceptual domains.

4. Discussion

We have looked at many examples of motor system involvement in perceptual processing, including speech perception, music perception, action perception, and non-human event perception. We have also discussed many of the existing theories of what role the motor system plays in perceptual processing, focusing on predictive theories, simulation theories, and the theory of neural reuse. All of these accounts and explanations vary widely in scope and in content, but as we have seen, there is significant overlap among them., which I will highlight here.

The mirror neuron theory proposes motor involvement due to the mapping of action to observation that is carried out in special neurons whose purpose is to understand action. This could give rise to many of the cases of motor activity in perception that we discussed here. Mirror neurons could underlie the mapping between a speech sound and its corresponding motor command; they could also perform the mapping of a musical sound, such as a piano piece, with the movements necessary to make that sound. This theory does not, however, explain motor activation during beat perception (there is no required motor command for this) or during non-biological object perception. Jeanerrod's simulation theory falls short here as well.

The emulator/prediction theories suggest that perception occurs via online prediction and modeling of the observed system and its regularities, such that each next state of the system is predicted. These theories are similar to simulation theories, but don't assume an ongoing simulation of the observed system using one's own body. Rather, the focus is on internalizing the behaviors and regularities of that system for prediction of the upcoming state. This could potentially still result in simulation in contexts where it is useful. Emulator accounts can explain all of the above results, from music and speech perception to the perception of non-biological stimuli and beat perception. The parts of the motor system that were reportedly active during perception of the non-human stimuli consisted of PMC and pre-SMA. Areas active during beat perception include PMC, dPMC, SMA, pre-SMA, and cerebellum. vPMC was the only brain region that showed activation during *only* human-created stimuli in the reviewed experiments.

In the reviewed neuroscience literature, a clear role for any particular region of the motor system does not emerge. Different experiments find activation in different regions for what they state is the same sort of task. At this point, the picture looks like each of the motor regions is involved in many, if not all, of the observed tasks. As we accumulate more theories and experimental results in this area, the particular role of the motor system in each of these instances of perception will be made clearer. Identifying the commonalities between specific cognitive and perceptual processes will lead us further toward understanding some potential basic non-motor functions of what is likely more than just the "motor system".

Chapter 2
Corticospinal excitability during the processing of handwritten and typed words and non-words.

A number of studies have suggested that perception of actions is accompanied by motor simulation of those actions. To further explore this proposal, we applied Transcranial magnetic stimulation (TMS) to the left primary motor cortex during the observation of handwritten and typed language stimuli, including words and non-word consonant clusters. We recorded motor-evoked potentials (MEPs) from the right first dorsal interosseous (FDI) muscle to measure cortico-spinal excitability during written text perception. We observed a facilitation in MEPs for handwritten stimuli, regardless of whether the stimuli were words or non-words, suggesting potential motor simulation during observation. We did not observe a similar facilitation for the typed stimuli, suggesting that motor simulation was not occurring during observation of typed text. By demonstrating potential simulation of written language text during observation, these findings add to a growing literature suggesting that the motor system plays a strong role in the perception of written language.

Introduction

Language is a deeply embodied system. We speak using our tongue and mouth muscles, we write using our hands, and we learn the meanings of words by observing the sensory and motor features present while hearing those words. Understanding the role that motor activation plays in each context of language processing is an ongoing enterprise. Many processes considered to be a part of the motor system have been revealed to have involvement in language (Fadiga et al., 2002; Fischer, Zwaan, & Fischer, 2017; Glenberg & Kaschak, 2002; Pulvermuller, 2005). Several competing explanations exist as to why non-motor cognition and perception would call on the motor system, including simulation theories (Barsalou, 2009; Gallese, 2007; Gallese & Lakoff, 2005; Jeannerod, 2001), active prediction theories (Clark, 2015; Lupyan & Clark, 2015; Schubotz, 2007), and motor resonance theories (Iacoboni, 1999).

A large body of work has looked into understanding the relationship between the motor system and language use in humans (Pulvermuller, 2005; Candidi et al, 2010; Hauk et al., 2004; Oliveri et al., 2004; Yang & Shu, 2006. One theory called the “motor theory of speech perception”, put forth by Liberman and Mattingly (1985), proposed that speech perception entails mapping the acoustic patterns of sound onto the gestures that are used in their creation. Fadiga et al. (2002) hypothesized that the mapping of these gestures involves mapping to their own respective motor system, in which case we should see activation of the mouth motor region of someone listening to speech. They applied single-pulse transcranial magnetic stimulation (TMS) to the cortical tongue region of participants as they passively listened to words with either a double “rr” phoneme or the double “ff” phoneme. Motor-evoked potentials (MEPs) measuring cortico-spinal excitability were obtained from the tongue muscle using electromyography (EMG). Higher MEPs were observed in the “rr” condition, whose pronunciation involves more movement of the tongue muscle, suggesting that participants were in fact using their own motor regions during speech perception. Skipper, Nusbaum, and Small (2005) found that

there was even greater increased motor activity while participants both saw and heard faces speaking, compared to only hearing or only seeing.

With the exception of Fadiga et al. and Skipper et al.'s findings, most of the research on the role of language in the motor cortex has focused on motor processing of action-based language, or the semantic content of language. Numerous studies, for example, provide evidence that action language, whether written or heard, and in words or full sentences, relies on the motor system for processing (Hauk et al., 2004; Oliveri et al., 2004; Tettamanti et al., 2005; Wilson, Saygin, & Sereno, 2004). However, language is also created using the motor system. As Fadiga et al.'s findings demonstrate, hearing spoken language relies on the activation of the mouth region of the cortical motor system.

Written language has been less explored in the context of the motor system. We learn reading and writing using our sensorimotor system to write letters and words on paper or type them on a keyboard. A recent behavioral study by Beilock and Holt (2007) found evidence that skilled typists may be simulating typed letters as they perceive them. They asked participants who were either expert or novice typists in an experiment to choose which of two competing letter dyads they liked better. Participants chose between a dyad of two letters that require the same finger using traditional typing methods [i.e., F, V] or a dyad of two letters that require different fingers using traditional typing methods [i.e., F, J]. They found that experts had a slight preference for the dyads that used different fingers to produce each letter, while novices did not exhibit a preference for either option. A motor task performed while making dyad preference judgments attenuated the preference of the skilled typists but only when the motor task involved the specific fingers that would be used to type the dyads. These findings suggest that in skilled typists, perceiving letters involves sensorimotor simulation of typing, which in turn influences affective judgments such as likeability.

In line with the abovementioned results, we designed an experiment to measure activation of the motor system during the perception of written language. For this purpose, we applied single-pulse TMS over left M1 and recorded MEPs from the right first dorsal interosseous (FDI) muscle in the right hand while participants saw words or non-words typed out or handwritten. We used only non-action words to avoid the recruitment of the motor system for the semantic component of action language. We predicted that during the appearance of typed or handwritten text, simulation of an inferred agent typing or writing would cause an increase in corticospinal excitability measured by MEPs. The motivation behind this experiment was twofold. The major aim was to extend theories of language embodiment to written language. We also aimed to further our understanding of the role of the motor system in non-motor processes such as language perception. While the present experiment was not aimed to distinguish between any existing theories of motor involvement, testing action observation in more and different contexts can add to this emerging area of research.

Methods

Participants

Twenty-four right-handed normal participants (8 males, 16 females, mean age ~19.5) were recruited in this study through UC Merced's SONA research system. All

participants passed a safety screen and gave written, informed consent. The experimental procedure was approved by the UC Merced Institutional Review Board for research ethics. Participants received 2 research credits that can be used for credit in some undergraduate courses.

TMS and EMG recording

Corticospinal excitability was measured by the amplitude of motor evoked potentials (MEPs) recorded using electromyography (EMG) on the first dorsal interosseous (FDI) muscle of the right hand. MEPs were chosen as the primary measurement because we were targeting corticospinal excitability during passive observation while subjects rested their hand. Related measures also reported in the literature, such as cortical silent period or MEP recruitment curves, could provide a more detailed measure of corticospinal excitability. However, due to constraints on number of stimulations we wanted to apply to participants and the desire for passive observation, MEP amplitude was the optimal measure for our purposes. Two small adhesive electrodes (1cm2ˆ) were placed over the belly of the recorded muscle and a ground electrode was placed over a bone on the participant's elbow. A bandpass filter (50 Hz–1000 Hz) was applied to the EMG signal, which was digitized at 1000 Hz for offline analysis. MEPs were elicited by applying single-pulse TMS to the FDI region of the left motor cortex. Pulses were delivered using a Magstim Rapid2 TM with an attached 70 mm figure-of-eight coil positioned over the optimal scalp location with the handle pointing backward at 45 ◦ from the midline. The procedure was as follows. Subjects were fitted with a swim cap that was covered by a grid of dots placed 1 cm2 apart. Optimal scalp position was determined by moving the coil by one centimeter intervals until the location eliciting the best MEPs was identified. This location was marked on the swim cap worn by the participant. After determining the stimulation site, we relied on VisorTM (ANT-Neuro Enschede, Netherlands) − a motion capture based neuronavigation software to ensure that the coil does not move during the duration of the experiment. This method allows for accurate repositioning throughout the experimental sessions and is consistent with the standard methods used for stimulation of M1. Resting motor threshold was determined as the percent of machine output that produced 5 out of 10 MEPs of at least 50 mV peak-to-peak amplitude. The methods described here are very similar to our previous work involving stimulation of the primary motor cortex (Therrien, Richardson, & Balasubramaniam, 2011; Therrien, Lyon, & Balasubramaniam, 2013). The stimulation intensity during the experiment was set to 120% of a participant's resting motor threshold. The coil was held steady at the optimal position throughout the experiment. Subjects were instructed to keep their head still and remain relaxed for the duration of the experiment.

Experimental paradigm

The visual stimuli consisted of videos of either handwritten or typed words or non-words appearing letter by letter at a variable presentation speed averaging 3–4 letters per second. Non-words in this experiment were groups of consonants. Words and non-words were the same length (between 6 and 8 letters). Words were chosen that did not relate to any actions or manipulable objects, to ensure that our measurement would not be influenced by the effects of semantic processing of action. We also included 10 baseline

trials, which consisted of a single black box for the same duration as the stimuli. We chose to randomize the baseline trials in with the rest of the trials so that the baseline measure would not be biased by a lack of attention that can occur when baseline measures are all recorded pre-experiment. Stimuli included five words and five non-words, which appeared four times in each of the conditions. This resulted in 80 stimulus trials and 10 baseline trials, or a total of 90 trials. Eight seconds passed in between individual trials, and the total experiment length was approximately 12 min. Because TMS stimulation would occur two seconds into the video, we ensured that the typed stimuli would display one of the following letters at that time [N, H, U, M, J, I], so that if subjects were simulating the typing, FDI would be the simulating muscle.

The stimuli appeared on a flat screen monitor placed in front of the participants. Participants were instructed to attend to the stimuli on the screen and were given notice when the experiment was one-third and two-thirds of the way finished to prevent loss of attention. Breaks were provided upon participant request. TMS pulses were delivered 2 s after video onset. The interval between trials was 8 s, to avoid any cumulative effects of single-pulse TMS. After the experiment, subjects were asked whether they were able to stay attentive during the length of the experiment. Participants who said they were not were excluded from analyses (5 subjects). The stimulus details are shown in Fig. 1 (please also see supplementary materials for video presentations).

Words		Nonwords	
Typed	Handwritten	Typed	Handwritten
volume	volume	hbvjmq	hbvjmq
skinny	Skinny	ycrbjt	ycrbjt
autumn	autumn	rqlhnf	rqlhnf
refund	refund	mjknmw	mjknmw
amount	amount	ptthnp	ptthnp

Fig.1. Sample stimuli used in the experiment. The typed stimuli appeared letter by letter. The handwritten stimuli appeared as if written out continuously. Videos used in the experiment are included in supplementary materials.

Results

In order to use inter-individual comparisons, Z scores were calculated for each participant. Trials in which MEP amplitudes were larger than 2.5 standard deviations from the mean and those less than 50 mV were excluded as outliers. Less than 5% of all data were excluded. Statistical analyses were carried out in R. Repeated measures two-way analysis of variance (ANOVA) was conducted on the normalized data for condition (typed or handwritten) and stimulus type (word or non-word). Fig. 2 shows the average Z-scores for each of the conditions.

As seen in Fig. 2, we found a significant main effect of condition ($F(1,23)=9.52$ $p<.01$), indicating that MEP amplitude was modulated by whether the stimuli were handwritten or typed. Specifically, the handwritten stimuli showed a much greater facilitation of MEPs than the typed stimuli.

The main effect for stimulus type (word or non-word) was not significant ($F(1,23) = .25$, $p > .6$), suggesting that motor cortex activation was not modulated by whether the stimulus was a real English word or a non-pronounceable consonant group. The interaction between condition and stimulus type also did not approach significance ($F(1,92) = .08$, $p > .7$), suggesting that the handwritten stimuli facilitation did not vary between words and non-words.

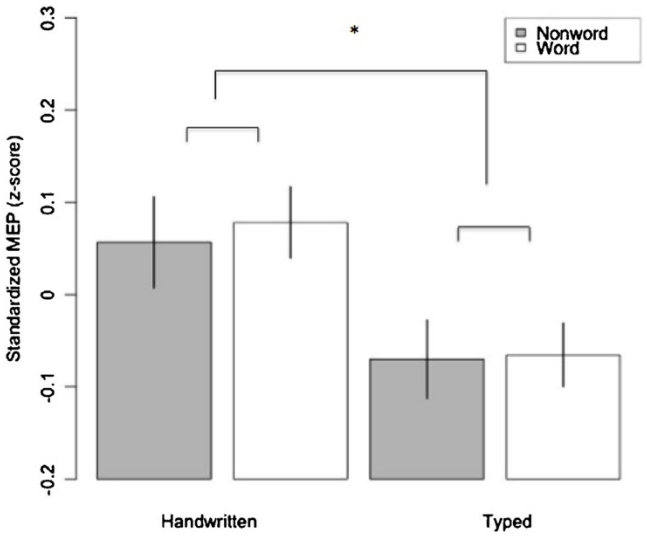


Fig. 2. Standardized (Z-scored) MEP amplitudes for each condition. Data from all subjects. Motor evoked potentials in the handwritten condition show facilitation. Vertical bars denote standard error means. Asterisks denote significant ($p<.05$) differences between conditions.

Discussion

In the present experiment, we found evidence for simulation of handwritten text during observation, regardless of whether the text segments were real words or groups of consonants. We did not, however, find evidence for simulation of typed text of the same nature. While the present experiment was not aimed to distinguish between any existing theories of motor involvement, testing action observation in more and different contexts can add to the evolving data that exists.

Here we show that passively observing handwritten words leads to an automatic facilitation of the reader's motor cortex. This automatic facilitation during reading perception is very similar to that found in Fadiga et al.'s speech perception experiment, where spoken stimuli involving greater tongue motion produce facilitation in MEPs recorded from the tongue muscle. An interesting difference in the present work is that if subjects are simulating an observed agent, in this case they must also infer an agent that is not present. In the case of our stimuli, this would mean that subjects are simulating the creation of the stimuli from a temporally-removed agent that previously created them. Evidence in favor of the simulation of inferred agents comes from some work in the action observation literature. Umiltà et al. (2001) found, during single-cell recording, that some subset of neurons in the macaque fire during the final part of an observed action, even if that final part of the action is occluded from view. Importantly, this suggests that

these neurons are simulating the action of an inferred agent when the actor is no longer in sight. Further evidence for this comes from work by Kohler and colleagues (2002), where they were also recording from single neurons in monkey premotor cortex. They found that some of the same neurons that fire during a produced and observed action will also fire when monkeys are only hearing the auditory information from the action (i.e., the cracking of a peanut). When only hearing the action, subjects must be inferring an agent.

There are other potential explanations for modulation of MEPs in the handwritten stimuli. One possibility is that participants are simulating writing the stimuli themselves, without inference to another agent. An interesting follow up in this regard would be to observe how MEP amplitude changes if handwritten words appear in a participant's own handwriting, or if MEPs are measured on the non-dominant hand during TMS of the contralateral motor cortex. Another possibility is that the motor system is active in sensory prediction of the motion of the handwritten stimuli. More specifically, the motor system might be using something that Wilson and Knoblich (2005) refer to as emulators, whereby perceptual prediction of the very next sensory state of a stimulus is being modeled using the motor system. While we might expect that we would see the same modulation in the typed stimuli if prediction were responsible, perhaps the one-by-one appearance of typed text does not evoke the same kind of sensorimotor prediction as the continuous fluid motion of the handwritten stimuli. In other words, perhaps any continuously developing line would lead to activation of the motor system, whereas the instantaneous nature of all-at-once letter appearance does not lead to this motor recruitment. Recent work by Schubotz (2007) suggests that the motor system (premotor cortex in particular) is active during the perception of inanimate events, by showing that prediction of these events corresponds with activation in somatotopically-relevant areas of the premotor cortex. For instance, spatial prediction tasks lead to activation in regions of premotor cortex that typically share activation for executed/observed foot-related actions, whereas rhythm/pitch prediction tasks lead to activation in areas corresponding to executed/observed mouth actions. One way to explore this potential mechanism would be to measure corticospinal excitability during the appearance of a continuously developing line on screen and comparing to the presently obtained results.

Additionally, written language is learned in an embodied manner, learning letters and words via the process of using our motor system to create them. What our results suggest is that even the simple perceptual processes involved in reading handwritten language is embodying these learned motor reproductions of text. While it is interesting that this strong effect does not hold for the typewritten words, it is perhaps not that surprising. In a world where we read text from digital devices constantly, this connection between text and motor commands is not as direct and strong as that with handwritten language, except perhaps for expert typists (Beilock & Holt, 2007). As our society moves away from the use of handwriting and more toward text being produced primarily with technological means such as typing, though frequently with thumbs instead of fingers, it will be interesting to see whether this embodiment of language changes as a result.

On the methodological front, it would be of interest to measure modulation of corticospinal excitability using an active measure, such as cortical silent period (CSP), while subjects were actively contracting the relevant muscle, for instance by holding their hand in a position primed for handwriting. However, the large body of work in this area

including work on action observation (Fadiga et al., 2002; Iacoboni, 1999; Candidi et al., 2010; Skipper, Nusbaum, & Small, 2005) use the MEP as the dependent measure for quantifying cortico-spinal excitability. More work is required to compare the relationship between active CSPs and MEPs in tasks such as the one we have used in this paper. It would also be useful to see how active motor threshold used in repetitive stimulation studies (Therrien, Richardson, & Balasubramaniam, 2011; Therrien, Lyons, & Balasubramaniam, 2013) can be used as a dependent measure during action observation experiments. While the behavioral responses make mapping motor cortical locations straightforward, methodological improvements can be made by using navigation based stimulation for recording MEPs (Julkunen et al., 2009).

Our results add to a growing body of literature suggesting that recruitment of the motor system is widespread, even in contexts with less obvious action-related perceptual information. Language in particular is a multimodal embodied system, showing reliance on the motor system for spoken language, written language, and the understanding of semantics. We also add to the evidence in favor of embodied simulation by introducing another instance of embodiment, whereby the perceptual-cognitive process of reading handwritten text involves motor simulation. Moving forward, it is important to observe how motor recruitment changes with changes in the environment. We see that handwritten stimuli involves motor simulation when the actual writing is observed, but what about during observation of static handwritten text that was created beforehand? Future directions for this work include exploring how repetitive stimulation of TMS to create virtual lesions over important sensorimotor regions modifies language perception and the recruitment of the motor system for language. Other work aims to measure the potential additive effect of simulation of written text and of action words to see if motor activation is higher when both of these forms of language embodiment are present.

Chapter 3
Recruitment of the motor system during music listening: An ALE meta-analysis of fMRI data

Several neuroimaging studies have shown that listening to music activates brain regions that reside in the motor system, even when there is no overt movement. However, many of these studies report the activation of varying motor system areas that include the primary motor cortex, supplementary motor area, dorsal and ventral pre-motor areas and parietal regions. In order to examine what specific roles are played by various motor regions during music perception, we used activation likelihood estimation (ALE) to conduct a meta-analysis of neuroimaging literature on passive music listening. After extensive search of the literature, 42 studies were analyzed resulting in a total of 386 unique subjects contributing 694 activation foci in total. As suspected, auditory activations were found in the bilateral superior temporal gyrus, transverse temporal gyrus, insula, pyramis, bilateral precentral gyrus, and bilateral medial frontal gyrus. We also saw the widespread activation of motor networks including left and right lateral premotor cortex, right primary motor cortex, and the left cerebellum. These results suggest a central role of the motor system in music and rhythm perception. We discuss these findings in the context of the Action Simulation for Auditory Prediction (ASAP) model and other predictive coding accounts of brain function.

Introduction

In the case of (most) music, we do not merely passively receive temporal patterns, but actively engage with the sound stream by discerning an underlying periodicity. This profound shaping of temporal perception is central to understanding and participation in music, dance and even speech/conversation. In recent years, neuroimaging studies have shown that passively listening to music activates brain regions that reside in the motor system proper. The same neural correlates underlying the creation of music and moving to music appear to be involved even when one is only listening to a musical piece (Baumann et al., 2007; Meister et al, 2004; Bangert et al., 2006; Bengtsson et al., 2009; Chen, Penhune, & Zatorre, 2008; Grahn & Brett, 2007; Haueisen & Knosche, 2001).

The motor system has received increasing attention in non-purely-motor domains (Watkins, Strafella, & Paus, 2003; Liberman & Mattingly, 1985; Gallese & Lakoff, 2005; Schubotz, 2007; Fadiga, Craighero, & Olivier, 2005). Activity in motor regions during perception of human actions and language is ubiquitous. In early theories of cognitive processing, motor processes and perceptual processes were understood as entirely separate and encapsulated mechanisms (Fodor, 1985). As evidence accumulates that shows this is not the case and there is substantial overlap among the domains, theories of action production and action perception must be informed accordingly. More recent proposals argue for common coding of perceptual and motor information (Prinz, 1990; Hommel, 2013) that arises primarily due to the co-activation of perceptual and motor components of a given action. Anderson's (2015) theory of neural reuse additionally suggests that we should expect newly evolved functions such as language to make use of previously instantiated neural mechanisms whose computational functionality can be co-

adapted for new purposes. As such, it is likely that motor activation observed during speech perception, for instance, corresponds to a sharing of computational or functional resources for perception and production of a given speech sound. We can expect a similar sharing of resources for music production and music perception as well.

Patel and Iversen (2014) advanced a theory of motor activation during music perception called the ASAP (Action Simulation for Auditory Prediction) hypothesis. The theory suggests that the same neural underpinnings involved in the simulation of body movements are utilized by the motor planning system to entrain neural activation with musical beat. This entrainment allows simulations to be used online during music listening as a predictive sensory signal for the upcoming music beat. The simulation is not tied to a particular effector-based movement, but a simulation of a timed, rhythmic motion. Patel and Iversen suggest the dorsal auditory stream as a potential underlying neural pathway for this process (Rauschecker & Tian, 2000).

Rauschecker (2011) has also proposed that a unified function of the dorsal stream may be anticipatory control of sensorimotor events. In particular, he suggests the posterior superior temporal (ST) regions, along with the inferior parietal lobe (IPL), interface extensively with PMC, linking the motor and auditory regions to established sensorimotor networks for audiomotor processes, such as speech and music. This network is established through similar mechanisms to those used in motor control theories (Grush, 2004; Wolpert, Doya, & Kawato, 2003), where a feed-forward projection carrying an efference copy of a motor command is used as a prediction of the upcoming sensory consequences, which can then be compared with the actual sensory outcome of the motor act. Rauschecker proposes that the projection from inferior frontal gyrus (IFG) to vPMC is a likely candidate for carrying an efference copy, while IPL to posterior ST would carry an "afference" copy of the predicted motor signal, thus allowing a continuous audiomotor predictive loop underlying smooth perceptuomotor abilities.

Another candidate theory, suggested by Schubotz (2007), is of active sensory prediction of events using the motor system. Schubotz extends the idea of emulators from motor control theory to encompass inanimate event perception in addition to human action prediction. Schubotz suggests the following: when we repeatedly hear a melody, the lateral PMC builds up sensorimotor representations using input from association areas of the cortex. These sensorimotor representations are only audiomotor, lacking the proprioceptive-motor representations that are involved in an actual performed movement. The lateral PMC eventually establishes an internal model of this melody which can be used for perceptual prediction. This internal model is similar to that involved in motor control, but with the components for movement and sensory feedback removed. Schubotz proposes what she calls the HAPEM (Habitual Pragmatic Event Map) framework, which states that "the prediction of an event that is structured with regard to a property P engages the area of the lateral premotor cortex that is best adapted to specify its motor output in terms of property P". What this means is that the perception of events with different properties recruits particular somatotopic regions of vPMC, selected based on similarity to the underlying properties of that area of vPMC. For instance, the regions of vPMC that correspond to executing and observing mouth movements are recruited for the perception of rhythmic events, due to the underlying rhythmic nature of the vocal system.

The above theories all posit that cortical motor areas play a role in music listening. Another emerging theme is that of the motor system having a predictive role in perceptual processes. These accounts are primarily in agreement in terms of which sub-areas in the motor system are involved. Schubotz's framework directly proposes involvement of both lateral PMC and pre-SMA/SMA, while ASAP and Rauschecker's theory both set the dorsal auditory stream (which includes dPMC) as the primary substrate. However, activated regions within the motor system measured by neuroimaging methods tend to vary between research studies. For instance, numerous music listening experiments report motor activity in both supplemental motor area (SMA) and dorsal premotor cortex (dPMC) (Baumann et al., 2007; Meister et al, 2004; Bangert et al., 2006; Bengtsson et al., 2009; Chen, Penhune, & Zatorre, 2008; Grahn & Brett, 2007). Among these studies, a few show neural activations in cerebellum (Meister et al, 2004; Bangert et al., 2006; Bengtsson et al., 2009) or primary motor cortex (M1) (Haueisen & Knosche, 2001) during a music listening task. Said differently, most music-listening studies do not show activation in every region of the motor system, nor do they show uniform activation in any one part of the motor system. In order to gain insight into the functional contribution of the motor system to passive music perception, one necessary step is to determine which motor regions are consistently contributing across music listening instances.

There are many factors likely to contribute to differences across studies, as each individual experiment has its own musical stimuli that vary in terms of particular characteristics, such as rhythmicity, familiarity, and valence of the music, for instance. Stimuli consisting of highly regular rhythmic structure might engage brain regions important for timing and sequential structure (i.e., supplementary and pre-supplementary motor areas and the cerebellum), while others might not. Experiments also vary in terms of what a participant is directed to focus on in these paradigms, ranging from complete passive listening (not attending) to judging beat or other characteristics of the stimuli. Such task demands are also likely to influence which regions are active, as directing attention to a stimulus may encourage focusing on particular aspects of the music, such as its beat or rhythmicity. In the present study, we are interested in discovering what motor regions are engaged during all music perception—those activated during passive listening. We define passive listening as attentive listening while remaining still (i.e., not tapping along to the music).

Identifying which regions are active consistently across all music listening tasks would help gain insight into the underlying processes and hone existing theories. Many theories outlining the functional contribution of individual motor areas exist, which can be used to determine what particular function is being carried out in a task utilizing that motor region. If one critical component is the dorsal auditory stream, which has a proposed role in motor planning and mapping auditory information onto potential motor acts, we should observe observation in dPMC (Hickok & Poeppel, 2004; Hoshi & Tanji, 2007). If activation is found in vPMC, the underlying mechanism might be similar to that proposed in the action observation network, which is responsible for mirror system activity for observed and produced actions (Fadiga & Fogassi, 1995; Rizzolatti et al., 1996, Rizzolatti, Fogassi, & Gallese, 2001). Many studies that involve music with beat manipulation report activity in SMA and pre-SMA regions, which are presumed to be

important for sequential processing of action-related stimuli and for inhibition of movements, respectively (Halsband et al., 1993; Nachev, Kennard, & Husain, 2008) Thus, SMA activity might indicate processing of sequential aspects of the music, and pre-SMA the inhibition of the natural tendency to move or sway to the music. We also might observe activation of structures in the basal ganglia, which appear to be involved in beat perception (Grahn, 2009). The basal ganglia are important for movement timing and sequential movement execution (Zatorre, Chen, & Penhune, 2007; Harrington & Haaland, 1999). M1 activation corresponds to particular motor commands that are carried out by specific muscle groups (Grafton et al., 1991) and has also been found active during observation of actions (Hari et al., 1998; Porro et al., 1996; Jeannerod, 2001). Finally, the cerebellum is known for its crucial role in motor timing and coordination. Research on sensorimotor adaptation has long focused on the role of the cerebellum in predicting sensory consequences of movement and adapting to errors in these predictions (Marr, 1969; Martin et al., 1996; McDougle, Ivry & Taylor, 2016; Tseng et al., 2007). Furthermore, cerebellar activation in conjunction with hippocampal activity is thought to underlie spatiotemporal prediction of movements (Onuki et al., 2013). This implication in predictive processes of motor control might extend to imagined and simulated motor computations, e.g. the cerebellum might be active in musical prediction even when no direct motor control is required.

In order to determine which of these regions show reliable and consistent activation during music perception, we employed a meta-analysis of all neuroimaging experiments consisting of music listening using an activation likelihood estimation (ALE) (Turkeltaub et al., 2002). We predict that from this meta-analysis will emerge a pattern of activation that will enlighten and instruct future theories aiming to explain the motor-specific contributions to passive music perception. Activation of any of the motor regions will provide conclusive evidence for the involvement of the regions of the brain typically considered “action areas”, in the perceptual domain of passive music listening. This will inform theories about what roles are played by the traditional motor system.

Methods

Meta-analyses provide a formal, statistical integration to combine the results of several studies that address a set of related research questions. There are several methods available for the meta-analysis of neuroimaging data and careful consideration was given as to which was most appropriate for this study. First, our study aims were to synthesize neuroimaging data of studies comparing rest and passive listening. More specifically, we wanted to identify regions of consistent activation across studies. Activation likelihood estimation (ALE) meta-analysis (Turkeltaub et al., 2002) addresses this by treating the spatial relationship between within study foci as fixed effects and between study relationships as random effects. Secondly, we considered the characteristics of our dataset. Unlike some other methods (e.g., KDA and MKDA), ALE uses a Gaussian kernel. When several distinct foci are located within the same general area, the Gaussian kernel is most likely to recover the separate foci. And, in general, if the spatial error on peak locations is approximately Gaussian (a reasonable assumption), then the Gaussian kernel will likely yield the most sensitive results. To investigate our research questions, we conducted ALE meta-analysis. Imaging studies commonly report brain locations of

task-induced activations as coordinates in 3D space (x,y, and z). ALE meta-analysis techniques can be used to identify reliable activation patterns in 3D space across studies. ALE is a coordinate-based approach to a meta-analysis, allowing researchers to integrate imaging data. Studies are collected, coded and interpreted using analytical methods to assess likelihood of activation through agreement or overlap in activation patterns.

To perform the ALE meta-analysis, we began by first locating relevant studies. Relevant studies were those that utilized functional brain imaging of healthy subjects listening tasks. We conducted literature searches in Medline and the BrainMap database (Laird, Lancaster & Fox, 2005) using a combination of the following: (1) a functional brain imaging modality, including positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) and (2) relevant adjectives related to auditory stimuli. For example, a single search consisted of "Imaging" AND "passive listening" OR "fMRI or functional magnetic resonance imaging" AND "auditory". The literature search of Medline was performed February 2016 and returned 132,294 papers. The literature search of BrainMap was performed September 14, 2016 and returned 244 studies. To ensure our ability to investigate the specified research questions a subsequent study selection process was done by applying the following inclusion criteria to the studies: (1) subjects were healthy adult participants; (2) The analyses include contrasts against rest or a suitable low-level control condition; (3) peak coordinates of group-level activations were reported; (4) foci activation were available in the form of standardized stereotaxic coordinates in either Talairach or Montreal Neurological Institute (MNI) space; (5) that results from the entire scanned volume were reported; and (6) data were available as of September 2016. An effort was made to obtain unreported coordinates from selected studies meeting all other criteria, however, this effort did not return any results. The subsequent review process was performed in two phases. First, an automated review of study titles was done using the R environment (R Development Core Team, 2008) to remove studies that were not in healthy human subject populations. The automated review removed 8144 papers from the database. Next, reviewers read the abstract and/or methods sections of remaining studies to assess appropriateness using the above inclusion criteria. Fig 1 illustrates the full review process for the meta-analysis. The process yielded 42 experiments that met the criteria for inclusion. A full list of experiments included can be found in Table 1. Experiments included a total of 386 unique subjects, approximately 195 male and 171 female.

Coordinates (X, Y, Z) for selected studies were recorded and, where necessary, transformed to Talairach space. Coordinates from individual studies were transferred to a text file formatted for analysis in GingerALE 2.3.6 (http://www.brainmap.org/ale/; Research Imaging Center, University of Texas, San Antonio, TX). These were transferred either using brainmap's Sleuth software (if the studies were located in the brainmap database), which outputs coordinates in the correct format for GingerALE, or were transferred individually by hand. The ALE meta-analysis was carried out in GingerALE. The ALE procedure was as follows: (1) model of single-study activation foci as peaks of three-dimensional Gaussian probability densities with subject-based full-width at half-maximum values (Eickhoff et al., 2009); (2) summation of probability densities to produce a statistical map estimating the likelihood of activation at each voxel; (3) thresholding of this ALE map based on the null hypothesis of a uniform distribution of

foci; (4) correcting for multiple comparisons by family-wise error thresholding. Resulting statistical maps show clusters where convergence between foci is greater than would be expected by chance. Statistical maps were thresholded using cluster-level family-wise error correction P<0.05 (cluster-forming threshold voxel-level P<0.001).

We split the data into separate studies that used either musicians only or nonmusicians only, with the intention of performing a contrast analysis between the two groups. Unfortunately, there were too few studies in these groups individually (14 experiments in each group), so we were unable to complete this contrast.

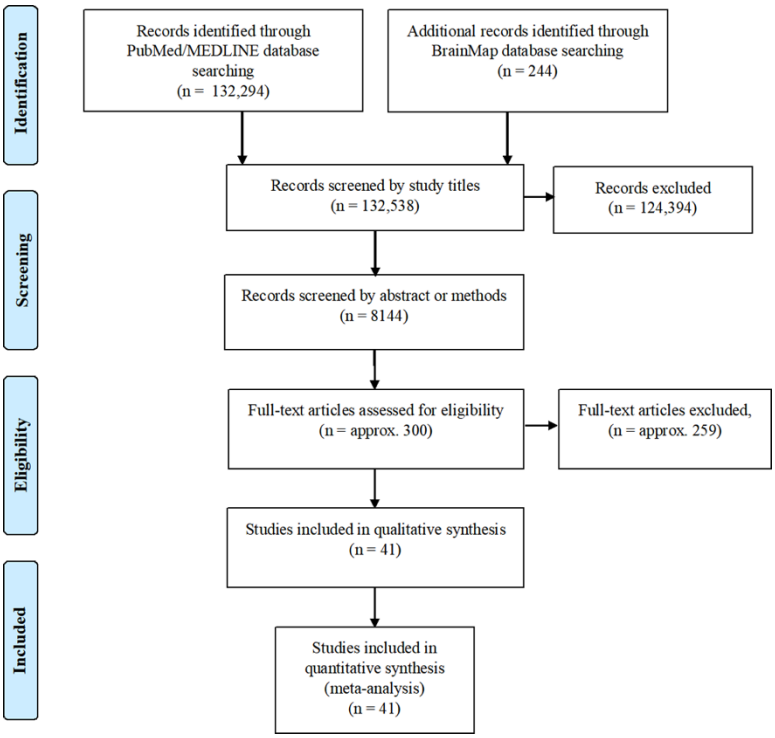


Fig 1. Flow diagram of study review.

Table 1. A list of the studies and experiments that were part of our meta-analysis.

Experiment	*Subj*	*Foci*	*Comparison*	*Instructions*	*Music Type*	*Musical Training*	*Age*	*Male*	*Handedness*
Alluri et al., 2012 [54]	11	215	Listen vs. rest	Remain still and to relax while listening to the musical stimulus and to maintain their gaze on the screen	Modern tango (*Adios Nonino by Astor Piazzolla*)	Mean years of music training 16.1 ± 6 SD	Mean age 23.2 ± 3.7 SD	6 (55%)	Unknown
Baumgartner et al., 2006 [55]	9	27	Listen and look at picture vs.	Instructed the subjects to place	Emotional classical	Unknown	Mean age 24.8; range 21-30	0 (0%)	Right

Experiment	*Subj*	*Foci*	*Comparison*	*Instructions*	*Music Type*	*Musical Training*	*Age*	*Male*	*Handedness*
			fixation baseline	themselves into the same mood as expressed by the presented emotional stimuli	orchestra music				
Blood et al., 1999 [56]	10	4	Listen vs. baseline (acoustically matched noise bursts)	Instructed to listen carefully. After the scan, subjects used a bipolar rating scale to rate emotional valence and intensity of stimuli.	Novel emotional music with varying dissonance	No more than amateur training	Unknown	5 (50%)	Right
Blood et al., 1999 [56]	10	8							
Brown et al., 2004 [57]	10	21	Listen vs. rest	Instructed to listen attentively to the music with their eyes closed without making any movement or response.	Wordless, instrumental rembetika style songs (unfamiliar to participants)	Nonmusicians	Mean age 33.8; range 21-51	5 (50%)	Right
Brown et al., 2007 [58]	11	57	Listen vs. rest; Listen and discrimination task vs. control (button press)	Melody listening: listen with eyes closed. Discrimination task: Listen and button press.	Piano melodies and harmonies, primarily adapted for this work.	University music education majors with a mean of 5.0 years of formal music instruction in voice or instrument. Having had an average of 12.3 years of involvement in musical production.	Mean age 24.6; range 19-46	5 (45%)	Right
Caria et al., 2011 [59]	14	20	Listen vs. silent control	Instructed to passively attend to music.	Instrumental pieces	Nonmusicians	Mean age 24.3 ± 3.02 SD	6 (43%)	Unknown
Chen et al., 2008; Exp 1 [5]	12	18	Listen with anticipation vs. silent baseline	Listened attentively	Rhythmic music	Nonmusicians	Mean age 23.83; range 20-32	6 (50%)	Right

Experiment	*Subj*	*Foci*	*Comparison*	*Instructions*	*Music Type*	*Musical Training*	*Age*	*Male*	*Handedness*
Chen et al., 2008; Exp 2 (A) [5]	12*	17	Listen with anticipation vs. silent baseline	Passively listen	Rhythmic music	Nonmusicians	Mean age 24; range 19-34	6 (50%)	Right
Chen et al., 2008; Exp 2 (B) [5]	12*	9	Passive Listen vs. silent baseline	Passively listen	Rhythmic music	Nonmusicians	Mean age 24; range 19-34	6 (50%)	Right
Demorest et al., 2010 [61]	16	15	Listen vs. rest	Listen, followed by memory test.	Three music examples from the Western classical tradition, three examples from the Turkish classical tradition and three examples from the Chinese classical tradition	<1 year of private music lessons and <3 years of ensemble (e.g., choir and orchestra) participation	Mean age 28.6 years with a range of 20.1–45.1 years	8 (50%)	Right
Dobek et al., 2014 [62]	12	33	Listen vs. baseline	Administered pain (thermal stimulation)	Self-selected by participants	Non-musicians	Range 18-40 years	0 (0%)	Unknown
Flores-Gutierrez et al., 2007 [64]	19	7**	Music – noise	Instructed to remain attentively focused on the auditory stimuli as their only task	Complex emotional musical pieces	No formal musical training	Mean age 25 (SD = 3.05)	11 (58%)	Right
Grahn et al., 2007 [6]	27	12	Music - rest	Instructed not to move any part of their body during presentation of the rhythms, followed by response given by button press to rhythm discrimination task	Rhythmic sequences	Fourteen out of 27 had musical training, defined as over 5 years of formal musical training and current regular musical activity and 13 had no musical training (reported no formal musical training or	Mean age 24.5; range 19-38	19 (70%)	Right

Experiment	*Subj*	*Foci*	*Comparison*	*Instructions*	*Music Type*	*Musical Training*	*Age*	*Male*	*Handedness*
						musical activities).			
Habermeyer et al., 2009 (A) [65]	16*	8	Listen vs. silent baseline	Watch silent movie without paying attention to the presented sounds	Deviant melodic patterns	8 trained lifelong musicians; 8 nonmusicians	Mean age 44.5 ± 9.9 years	14 (88%)	
Habermeyer et al., 2009 (B) [65]	16*	3	Listen vs. silent baseline	Watch silent movie without paying attention to the presented sounds	Standard melodic patterns	8 trained lifelong musicians; 8 nonmusicians	Mean age 44.5 ± 9.9 years	14 (88%)	Unknown
Heine et al., 2015 [67]	8	19	Music vs. baseline sounds	Instructed to keep their eyes closed, stay awake, avoid any structured thoughts, and listen attentively to the music	Dynamic musical excerpts chosen by loved ones from a list	Unknown	Mean age 26, SD ± 3	4 (50%)	Unknown
Hugdahl et. al., 1999 [68]	12	5	Musical instruments – simple tones	Button press at target sound	Excerpts from musical instruments	Unknown	Range 20 - 30	12 (100%)	Right
Langheim et al., 2002 [69]	6	4	Passive listening vs. Rest	Passive listening	Classical music (Vivaldi's Concerto in G minor, Bach's Suite in C major, part 2, Partita 2 and Partita 3)	At least 15 years of musical experience (two violinists, one pianist and three cellists); mean length of study 19.6 years, range 15–26 years	Mean age 27; range 22-32	2 (33%)	Right
Leaver et al., 2009; (A) [70]	10	9	Familiar and unfamiliar music	Subjects were instructed to attend to the stimulus being presented and to imagine, but not vocalize, the subsequent melody	Short piano melodies constructed for this experiment	At least 2 years musical experience (mean = 6.5, sd = 4.17)	Unknown	Unknown	Unknown
Leaver et al., 2009 (B) [70]	9	3	Familiar and unfamiliar music	Subjects were instructed	Short piano melodies constructed	Nonmusicians	Unknown	6 (67%)	Unknown

Experiment	*Subj*	*Foci*	*Comparison*	*Instructions*	*Music Type*	*Musical Training*	*Age*	*Male*	*Handedness*
				to attend to the stimulus being presented and to imagine, but not vocalize, the subsequent melody	for this experiment				
Mirz et al., 1999 [71]	5	7	Music – baseline	Subjects were asked to listen to the presented sounds without performing any semantic, phonological, temporal, intensity, or pitch analysis	Classical music (W.A. Mozart, Piano Concerto No. 21, 65 dB SPL)	Unknown	Mean age 34; range 24-50	2 (40%)	Right
Morrison et al., 2003 (A) [73]	6*	3	Music vs. rest	Following the scan subjects completed a poststudy recognition test	3 Baroque Western examples	Trained professional violinists and violists	mean age 38.3 years	2 (33%)	2 left handed, 4 right handed
Morrison et al., 2003 (B) [73]	6*	3	Music vs. rest	Following the scan subjects completed a poststudy recognition test	3 Chinese examples	Trained professional violinists and violists	mean age 38.3 years	2 (33%)	2 left handed, 4 right handed
Morrison et al., 2003 (C) [73]	6*	2	Music vs. rest	Following the scan subjects completed a poststudy recognition test	3 Baroque Western examples	Non-musicians	mean age 34.2 years	2 (33%)	Right
Morrison et al., 2003 (D) [73]	6*	2	Music vs. rest	Following the scan subjects completed a poststudy recognition test	3 Chinese examples	Non-musicians	mean age 34.2 years	2 (33%)	Right
Ohnishi et. al., 2001 (A) [75]	14	5	Music vs. rest	Instructed to passively listen to music	Italian concert BMV 989 by J.S. Bach	>12 years of 4–8 h of training per day) with AP (n = 10) or relative pitch (n = 4)	Range 20-27	2 (14%)	Right

Experiment	*Subj*	*Foci*	*Comparison*	*Instructions*	*Music Type*	*Musical Training*	*Age*	*Male*	*Handedness*
Ohnishi et. al., 2001 (B) [75]	14	4	Music vs. rest	Instructed to passively listen to music	Italian concert BMV 989 by J.S. Bach	Nonmusicians (no formal education musical and never played an instrument)	Range 21-27	2 (14%)	Right
Rogalsky et. al., 2011 [77]	20	5	Melodies vs. rest	Passive listening	Simple novel piano melodies	Twelve participants had some formal musical training (mean years of training = 3.5, range 0–8)	Mean age 22.6 years; range 18–31	9 (45%)	Right
Satoh et al., 2001 [78]	9*	8	Music (alto) vs. baseline	Subjects were asked to listening to and concentrate on the tone of the alto part of the harmony, and make a sign when they heard the tonic tone	3 fairly unknown motets; musical pieces of harmonious style with four vocal parts, composed by Anton Bruckner.	Musicians (music students)	Mean age 21.8 years; range 21–28	9 (100%)	Right
Satoh et al., 2001 [78]	9*	10	Music (harmony) vs. baseline	Subjects were asked to listen to the melody as a whole, and make a sign upon hearing the minor chord	3 fairly unknown motets; musical pieces of harmonious style with four vocal parts, composed by Anton Bruckner.	Musicians (music students)	Mean age 21.8 years; range 21–28	9 (100%)	Right
Satoh et. al., 2003 [79]	11*	7	Music (soprano) vs. baseline	Subjects were asked to listen to the soprano part of the harmony, and make a sign when they regarded a tonal sequence as one phrase	Three new musical pieces of harmonious style with three vocal parts	Nonmusicians (no formal musical education or training)	Mean age 21.2 years; range 20–30	11 (100%)	Right

Experiment	*Subj*	*Foci*	*Comparison*	*Instructions*	*Music Type*	*Musical Training*	*Age*	*Male*	*Handedness*
Satoh et. al., 2003 [79]	11*	10	Music (harmony) vs. baseline	Subjects were asked to listen to the melody as a whole, and make a sign upon hearing a dissonant chord	Three new musical pieces of harmonious style with three vocal parts	Nonmusicians (no formal musical education or training)	Mean age 21.2 years; range 20–30	11 (100%)	Right
Satoh et. al., 2006 [80]	10*	16	Music (familiarity) vs. baseline	Subjects were asked to listen to the melodies and then judge whether the melody was familiar	33 melodies (27 melodies were well-known old Japanese nursery songs)	Nonmusicians (no formal musical education or training)	Mean age 21.6; range 20-28	10 (100%)	Right
Satoh et. al., 2006 [80]	10*	13	Music (alteration-detecting task) vs. baseline	Subjects were asked to listen to the same melodies and detect the altered notes by making a sign	33 melodies (27 melodies were well-known old Japanese nursery songs)	Nonmusicians (no formal musical education or training)	Mean age 21.6; range 20-28	10 (100%)	Right
Schmithorst, 2005 [81]	15	30	Melodies - random tones	Passive listening	30 s of an unharmonized popular melody, followed by 30 s of tones of random frequency and duration, followed by 30 s of the previous melody, harmonized using triads an octave below	7 out of 15 received prior formal musical training, receiving formal instruction, continuously from early childhood (8 years old) throughout adolescence	Mean age 37.8 ± 15.2 SD	11 (73%)	Unknown
Toiviainen et al., 2014 [82]	15	38			Comprised the B-side of the album Abbey Road by The Beatles (1969).	Unknown	Mean age 25.7 ± 5.2 SD	10 (67%)	Right
Trost et al., 2011 [83]	15	20	Music vs. random tones	Subjects were asked to listen closely and provided a rating of emotional feeling	Emotional classical music	No professional music experience	Mean age 28.8 +- 9.9	8 (53%)	Right

Experiment	*Subj*	*Foci*	*Comparison*	*Instructions*	*Music Type*	*Musical Training*	*Age*	*Male*	*Handedness*
				following the music piece					
Tsai et al., 2010 [84]	12*	7	Music - baseline	Subjects were asked to passively listen to unlearned percussion music	Sichuan opera percussion music, Beijing opera percussion music, syllable representation of Beijing opera percussion music, and Taiwanese opera tunes played by the erhu	Music training for more than 4 years	Range: 20-26	2 (17%)	Right
Tsai et al., 2010 [84]	12*	7	Music - baseline noise	Subjects were asked to listen and hum covertly along to learned percussion music	Sichuan opera percussion music, Beijing opera percussion music, syllable representation of Beijing opera percussion music, and Taiwanese opera tunes played by the erhu	Music training for more than 4 years	Range: 20-26	2 (17%)	Right
Tsai et al., 2010 [84]	12*	7	Music - baseline noise	Subjects were asked to listen and hum covertly along to the verbalized syllable representation of learned percussion music	Sichuan opera percussion music, Beijing opera percussion music, syllable representation of Beijing opera percussion music, and Taiwanese opera tunes played by the erhu	Music training for more than 4 years	Range: 20-26	2 (17%)	Right
Tsai et al., 2010 [84]	12*	7	Music - baseline noise	Subjects were asked to listen and hum covertly along to the verbalized syllable	Sichuan opera percussion music, Beijing opera percussion	Music training for more than 4 years	Range: 20-26	2 (17%)	Right

Experiment	*Subj*	*Foci*	*Comparison*	*Instructions*	*Music Type*	*Musical Training*	*Age*	*Male*	*Handedness*
				representation of learned melodic music	music, syllable representation of Beijing opera percussion music, and Taiwanese opera tunes played by the erhu				

* The same pool of participants was used for separate analysis/study protocols. These were considered separate experiments for the purposes of this meta-analysis because analyses were performed separately and/or the dependent variable was altered between conditions.
** The published data was missing one z coordinate. An attempt was made to contact the authors, however, we were unable to obtain the missing information.

Results

Fig 2 shows the activations during passive listening, demonstrating the common brain network underlying music perception. Talairach coordinates for these ALE foci are presented in Table 2. Activations were seen in the bilateral superior temporal gyrus, transverse temporal gyrus, insula, pyramis, bilateral precentral gyrus, and bilateral medial frontal gyrus. As shown in Fig 2, there was activation in the left and right premotor cortex (BA 6), right primary motor cortex (BA 4), and the left cerebellum.

An inspection of Table 3 reveals that Cluster 1 is centered over the right primary auditory cortex, and spans from BA 22 and BA 41/42 (primary and secondary auditory cortices) in the right hemisphere to BA 6 (right premotor cortex). Likewise, in the left hemisphere, cluster 2 is centered over the left primary auditory cortex, and spans from BA 22 and BA 41/42 (primary and secondary auditory cortices) in the left hemisphere to BA 6 (left premotor cortex). Cluster 3 reveals motor system activation in the right hemisphere, centered in right premotor cortex and spanning from premotor to primary motor cortex. Finally, cluster 4 is located in the left cerebellum. Fig 2 depicts the activation patterns seen bilaterally for a range of *z* values.

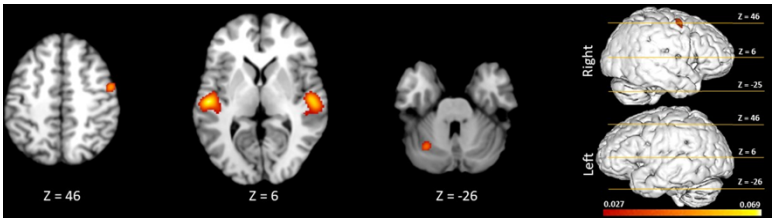


Fig 2. Significant clusters from meta-analysis of passive listening tasks in healthy volunteers (family-wise error correction (P<0.05)). The 3D brain is shown to indicate slice levels.

Table 2. Talairach coordinates for voxel clusters.

Area		BA	Conjunction			
			x	y	z	ALE
Temporal Lobe						
Superior Temporal Gyrus	R	22	52	-16	6	0.061329
	R	22	52	-6	-4	0.054407
	L	41	-52	-18	6	0.069281
	L	41	-42	-34	12	0.033931
Frontal Lobe						
Precentral Gyrus	R	4	50	-4	46	0.052819
Anterior Lobe						
Cerebellum (Culmen)	L		-28	-60	-26	0.045433

The Talairach coordinates of the significant ALE clusters are presented for the conjunction of passive listening ($p < 0.05$, FWE). The ALE values for the conjunction represent the minimum ALE value from the passive listening ALE maps. The ALE values shown are the true values times 10^{-3}. BA, Brodmann area.

https://doi.org/10.1371/journal.pone.0207213.t002

Discussion

We found evidence for consistent activation of various regions of the brain during passive music listening. As expected, our results showed activation in the primary and secondary auditory areas bilaterally. This is consistent with the existing literature showing that these areas are the critical regions of cortex for processing incoming auditory information (Schnupp, Nelken, & King, 2012). Other activated areas included both right primary motor cortex, right and left lateral premotor cortex, and left cerebellum. We discuss in turn the implications for each of these findings below.

Activation of premotor cortex

We were unable to pinpoint activation to any further subregions of lateral PMC (i.e., dorsal or ventral), as the activation pattern could be consistent with either dorsal or ventral PMC. The average coordinates for these regions overlap in such a way that neither can be ruled out. This means that premotor involvement could be via dorsal, ventral, or both. The activation of PMC in the present analysis is consistent with both ASAP and the HAPEM framework. However, because we do not know whether this activation is localized to ventral or dorsal PMC, it is unclear if this activity reflects involvement of the dorsal stream, or potentially the action observation network that recruits vPMC for action simulation. Also, given that these clusters only represent aggregate BOLD activation, we do not have insight into the temporal dynamics of this activity, which will be crucial for inference about its origin.

Activation of primary motor cortex

M1 activity could reflect either an excitatory or inhibitory contribution, as the BOLD signal does not differentiate the two. Vigneswaran et al. (2013) report that while many M1 neurons are active during action observation and thus classified as mirror neurons (Rizzolatti & Craighero, 2004), M1 neurons directly connecting to spinal circuitry and thus contributing to observed action are suppressed during the observation of action, in order to prevent explicit action. Simulative properties of mirror neurons have also been confirmed in response to auditory sounds (Kohler et al., 2002). Thus, either excitatory, inhibitory, or both could give rise to activation of M1 during passive music

listening. Some theories of mirror neuron activity (Friston, Mattout, & Kilner, 2011) additionally claim that the mirror neuron network uses active inference during the perception of observed actions using predictive mechanisms.

Further examination of the studies that contributed to the right premotor/primary motor activation cluster reveals that a number of them used tasks that were not passive listening in the same way as passive background listening. For example, Chen et al. (2008) required in some experimental conditions for participants to anticipate later tapping to the beat in subsequent trials, which may recruit motor planning regions during the listening task. Grahn and Brett (2007) asked their participants to determine whether the rhythms of two stimuli were the same or different, which may recruit motor areas to assist with the detection task. Finally, Tsai et al. (2010) asked participants in some of their tasks to covertly hum along with the music that they were hearing, which may recruit motor areas for subvocalization. Therefore, more work should be done to decisively conclude whether motor areas are recruiting during background listening, in addition to passively listening for properties of the music while remaining still.

Activation of cerebellum

Many studies contributed to the cluster indicating left cerebellum. The activation of PMC and cerebellum during music listening supports predictive theories of the motor system, such that the cerebellum might provide the predictive component in a forward model of the upcoming sensory consequences. The cerebellum may be providing an inverse model for mapping sensory input to the simulated movement that would give rise to that sensation. An investigation of the temporal dynamics of communication among these regions can again provide further insight into the mechanism.

Lack of activation in SMA/pre-SMA and basal ganglia structures

While quite a few studies did report activation of SMA and pre-SMA, we did not find corresponding activation in our meta-analysis. We also failed to find evidence of basal ganglia activation. One potential reason for this discrepancy may be that another process on top of passive listening is needed to engage these regions. The generally agreed upon roles of both SMA and the basal ganglia are of sequential learning and timing (Harrington & Haaland, 1999; Mushiake, Inase, & Tanji, 1990). Because these are properties of the majority of music, this region is likely to be recruited in many musical contexts. However, without directly listening for these properties of the music, it appears that automatic SMA and basal ganglia activation is not prevalent. Looking more closely at those experiments that explicitly report SMA/pre-SMA activation, they do appear to have a musical beat component to them, which relies on the underlying sequential and timing properties of the music. Bengtsson et al (2009) encouraged participants to focus on rhythmic properties of the music, as did Chen, Penhune and Zatorre (2008) and Grahn and Brett (2007). Baumann et al. (2007) required subjects to do a counting task during passive listening as a distractor, which may have resulted in SMA activation. Experiments showing basal ganglia activation also appear to involve a beat detection task [29].

It also may be the case that activation of SMA/pre-SMA is only prevalent in musically-trained individuals, who are more likely to attend to and perceive the structural

aspects of the music due to their background training. Baumann et al. (2007) report increased activation of both pre-SMA and SMA in musicians compared to nonmusicians, as did Bangert (2006) for SMA activation. Participants in Meister et at (2004) showing SMA activation were all musicians. Thus, it appears SMA activation is likely due to either a trained musical background and/or a focus on the rhythmic properties of music.

General Discussion

These results show that the passive perception of music engages a large and complex net- work of brain regions. This includes activation of areas in the motor system proper. Activation of the cerebellum and primary and premotor cortices suggests that perceived music is partly processed in areas typically considered as important for action-relevant information only. Recruitment of premotor areas during music listening supports many theories of motor involvement during perceptual tasks (Schubotz, 2007; Patel & Iversen, 2014; Rauschecker, 2011). The idea of shared neural resources for tasks with underlying computational similarities has gained recent theoretical and grounded neuro-biological support (Anderson, 2015). Most current theories suggest that perceptual processing involves the same kinds of temporal prediction involved in action, making a shared circuit useful for action-based and perception-based processing. An alternative (or potentially compatible) hypothesis is that involvement of PMC reflects the process of simulation (Jeannerod, 2001), where the motor system underlies simulation of the actions required to create the observed sensory information. Our findings are consistent with both of these theoretical frameworks, though it does not provide any insight for distinguishing which theory best fits the data, as this meta-analysis only tells us which areas are active at some point in the process. This work supports the currently merging conceptualizations of action and perception (Prinz, 1990; Friston, Mattout, & Kilner, 2011).

One limitation of the present meta-analysis was that we were unable to obtain data from contacted authors for studies that did not report all of the observed brain activations. It is possible that unpublished or unreported activations may have biased our results toward reporting motor and auditory areas, as studies that do not find activation in these areas of interest are less likely to be reported. This inability to obtain unavailable data also likely contributed to our inability to obtain enough studies for the musician/nonmusician contrast. Further exploration of differential activation in musicians relative to nonmusicians is important for advancing this work. Musicians exhibit plasticity-induced changes perceptual and motor abilities, as well as changes in structural and functional neuronal connectivity (Gaser & Schlaug, 2003; Hutchinson et al., 2003; Hyde et al., 2009; Kraus & Chandrasekaran, 2010). In particular, we believe that musicians passively listening to music should also recruit supplementary motor cortex, and might show greater activation of the cerebellum, which has a larger volume in musicians (Hyde et al., 2009). Another interesting avenue to pursue is to run more studies that directly compare different types of music listening tasks. For instance, we might compare background listening to listening in anticipation of some movement to listening for particular musical features, such as rhythm or grooviness. This will incorporate context-dependent music listening, which may reveal that different (but likely highly overlapping) networks are recruited in separate contexts. An additional limitation of this approach is that while we can identify which brain areas are active at some point during

the music listening task, the BOLD signal cannot tell us anything about the temporal dynamics of the process. Complementary methods, such as EEG, should be used along with imaging data to investigate the functional connectivity among these music listening networks. This will also allow us to determine which of the existing theories fit best with the data.

In summary, this study adds support to the idea that motor planning activity serves not only to help us move but is recruited for music perception even in the absence of movement. Further exploration will elucidate the functional purpose of this recruitment, as well as why and how different music listening contexts seem to engage slightly different networks. An understanding of the auditory-motor interactions underlying music perception could explain a growing number of findings suggesting an important link between music perception and the action systems of the brain.

Chapter 4
Affordance Compatibility Effect for Word Learning in Virtual Reality

Rich sensorimotor interaction facilitates language learning and is presumed to ground conceptual representations. Yet empirical support for early stages of embodied word learning is currently lacking. Finding evidence that sensorimotor interaction shapes learned linguistic representations would provide crucial support for embodied language theories. We developed a gamified word learning experiment in virtual reality in which participants learned the names of six novel objects by grasping and manipulating objects with either their left or right hand. Participants then completed a word-color match task in which they were tested on the same six words and objects. Participants were faster to respond to stimuli in the match task when the response hand was compatible with the hand used to interact with the named object, an effect we refer to as affordance compatibility. In two follow up experiments, we found that merely observing virtual hands interact with the objects was sufficient to acquire a smaller affordance compatibility effect, and we found that the compatibility effect was driven primarily by responses with a compatible hand and not by responses in a compatible spatial location. Our results support theoretical views of language which ground word representations in sensorimotor experiences, and they suggest promising future routes to explore the sensorimotor foundations of higher cognition through immersive virtual experiments.

1. Introduction

1.1. Action/sensory language is grounded in sensorimotor processes

Embodied language theories propose that linguistic representations are grounded in sensorimotor experiences: that words evoke sights, sounds, and movements in the mind, and those features constitute the representations of the words (e.g., Glenberg, 1997; Barsalou, 1999; Glenberg and Kaschak, 2002; Gallese & Lakoff, 2005; Barsalou, 2008). In many cases, theories of language embodiment have focused on how sensorimotor experiences are reactivated and recombined in mental simulations (Zwaan and Madden, 2005; Barsalou, 2009). Simulations are thought to constrain ongoing sensorimotor processes, perturbing the actions and perceptions of language listeners. Furthermore, interactions between the sensorimotor foundations of language and real time sensorimotor processes are bidirectional, thus concurrent motor and perceptual states bias the comprehension and production of language. Evidence supporting a general view of language as situated and embodied extends from embodied spatial language (Spivey et al., 2000) to emotional language (Glenberg, Havas, Becker, & Rinck, 2005; Havas, Glenberg, and Rinck, 2007) and abstract language (Casasanto & Boroditsky, 2008; Matlock et al., 2011).

In a seminal finding, Glenberg and Kaschak (2002) demonstrated how language processing can influence seemingly unrelated aspects of action. In an effect referred to as the action-sentence compatibility effect, participants heard sentences with implied motion

away from ("She closed the drawer") or toward a protagonist, and judged whether the sentences were sensible by pulling a lever toward or away from themselves. When the direction of the response was the same as the implied motion of the sentence, participants made sense of the sentences more quickly. The speed at which a response could be prepared and executed was influenced by the semantics of the sentence. This is consistent with a view that sentence meaning is understood by how, in terms of body and environment, the actions in the sentence are accomplished.

In addition to the substantial evidence indicating *that* language is grounded in sensorimotor experiences, numerous studies show *how* language generates patterns of activity in the nervous system which correspond to sensorimotor experiences. For instance, olfactory areas are activated by words associated with smells (González et al., 2006), and sound-related words like "ringing" activate auditory regions more than non-sound-related words (Kiefer et al., 2008). Many studies have shown that action words and action sentences activate somatotopic regions of the motor cortex, such that "kick" recruits leg area of vPMC and "lick" recruits the face area (Hauk, Johnsrude, and Pulvermüller, 2004; Tettamanti et al., 2005; Buccino et al., 2005), and this differential activation has been found to occur as early as 200 ms after word onset (Hauk and Pulvermüller, 2004), suggesting involvement of these areas in early semantic processing. Not only does language perception evoke sensorimotor activity, non-invasive brain stimulation of action regions can facilitate or inhibit linguistic processes. Pulvermüller et al. (2005) found that priming the respective effector representation area of primary motor cortex (M1) using single-pulse transcranial magnetic stimulation (TMS) decreased reaction times for responding to words describing actions performed by the stimulated effector. Vukovic et al. (2017) found that online repetitive TMS to the motor cortex slowed reaction times to action words, while leaving reactions to abstract words unaffected. These experiments bolster the view that sensorimotor systems in the brain are not only activated by language comprehension but play active roles in understanding.

1.2. How do novel words become grounded in sensorimotor systems?

Despite the abundant evidence that language is grounded in the experiences of the body, a key area of embodied language theories lacking empirical support is the process by which sensorimotor experiences come to underlie the representation of novel words. The studies described above observe neural activation during perception of well-known words, which reflects the long-term semantic networks of these concepts and top-down knowledge of affordances. Research on the acquisition of embodied language effects is sparse (see Richter, Zwaan, & Hoever, 2009; Öttl, Dudschig, and Kaup, 2017), and it has yet to be demonstrated that natural interactions with novel objects can give rise to the kinds of effects discussed above. Empirical support for this initial phase of embodied language learning provides a crucial test of embodied language theories. If novel words do not show early effects of the sensorimotor context in which they were learned, this would undermine the view that the sensorimotor processes are truly constitutive of the word meanings, rather than more passive associations. However, if specific sensorimotor experiences that take place during word learning influence how those words subsequently affect behavior, it would provide powerful support for grounded word learning.

Circumstantial evidence for sensorimotor interaction playing an important role in the acquisition of new words can be found in studies of infant word learning. Yu, Smith, and Pereira (2008) found in a novel word learning study with 18-month old children that the proportion of time that an object remained in an infant's visual field, as well as the amount of time holding a named object when its name was spoken, was predictive of successful word learning. This suggests that sensorimotor properties are important features of word learning, but leaves unclear whether the kinds of interactions experienced by a learner influence the semantic representations of learned words.

One reason that evidence for the acquisition of embodied language effects is sparse is due to the tendency for word learning studies to rely on standard computer tasks where participants learn the novel words for flat images of objects on a screen. This kind of learning is very unlike real-world word learning, where children learn about objects by picking them up and interacting with them. Research with real objects finds that real world objects are remembered more accurately than their photographic counterparts (Snow et al., 2014). Neuroimaging studies also show that the neural mechanisms involved in processing 3D objects may be distinct from mechanisms involved in processing 2D versions of those same objects (Snow et al., 2011). In addition, when action animations that align with the meaning of learned verbs are presented with the verbs, learning is greater than when those animations do not align with the verbs (Hald et al., 2015). This suggests that having concurrent representations of the actions implied by words improves learning of the words. This corresponds to how words are learned in the world, where a spoken word often co-occurs with the object or action it refers to, or a gesture indicating the action. This body of work suggests that realistic objects and movements will be more likely to result in embodied language effects.

We conducted a series of three experiments to investigate sensorimotor grounding of novel words acquired through sensorimotor interaction with objects in a virtual environment. In the first experiment, we investigated whether participants would be faster to respond to novel words that were learned through sensorimotor interaction when the action required for the response used the same hand and movement as the affordance learned for the word. We refer to this relationship as an affordance compatibility effect. One previous study has shown evidence of spatial congruency effects for novel words (Öttl, Dudschig, and Kaup, 2017). The authors had participants learn the names of novel objects in the environment in front of them that were either located in the upper or lower visual field. In a test phase, recollection of the objects was facilitated when participants made an up or down movement congruent with the original location of the object. This study reveals a spatial component of the learned representations; however it does not involve the kind of realistic sensorimotor interaction with objects thought to underpin natural word learning. In contrast, participants in our experiments learned novel object names in a virtual environment with naturalistic affordances and were then tested in a word-color match test. We found that participants acquired an affordance compatibility effect where they were faster to respond to matches in the test phase with the hand used to interact with the named object from the training phase. In a follow-up experiment, participants learned the same words by observing virtual hands interacting with the objects. This was done to investigate whether the affordance compatibility effect is dependent upon direct object manipulation. Finally, in a third experiment, we explored

the extent to which spatial affordance compatibility could be separated from effector-specific affordance compatibility.

2. Experiment 1: Direct manipulation induces affordance compatibility

In a first experiment, we explored whether sensorimotor experience during novel word learning would influence later processing of the learned words. We developed a gamified virtual reality experiment using Unreal Engine 4. Using virtual reality enabled us to attach specific manual affordances to virtual objects with rich visual properties. We predicted that learners would associate the affordances of objects with the words for those objects and that these associations would influence behavioral responses even when explicit retrieval of the affordances was not necessary. We tested this by comparing responses in a word-color matching task. If responses which were compatible with the affordance of a word were faster than incompatible responses, this would be evidence of a learned affordance compatibility effect. In addition, we incorporated a variety of visual and auditory consequences of actions (e.g. potions pouring a stream of liquid) to motivate learning and encourage participants to engage with the virtual environment.

Fig. 1: The HTC Vive virtual reality system consisting of a motion-tracked head-mounted display with 2160x1200 resolution, 2 handheld motion controllers, and 2 wall-mounted infrared sensors. During the experiment, participants cannot see their actual body or surroundings, but instead see a game-like environment and virtual hands (see **Fig. 2**). Image source: HTC Vive Press Kit.

2.1. Methods

Twenty-seven participants (23 women; 25 right handed) completed a two-part experiment using an HTC Vive virtual reality system (**Fig. 1**). We initially planned 30 participants, but due to technical issues and low performance during training, only 27 completed the experiment. Participants were adult undergraduate students (age 19-23 years) recruited from the University of California, Merced behavioral subjects research pool. All participants had normal or corrected vision and normal hearing and spoke fluent

English (19 bilingual). Participants provided informed consent prior to beginning the experiment. During the experiment, participants interacted via hand-held controllers with a virtual environment. The controllers were visually represented to the participants as virtual hands which tracked the position and orientation of the participant's actual hands. All participants completed a pre- and post-exposure comfort survey (supplementary materials) adapted from the Simulator Sickness Questionnaire (Kennedy, Lane, Berbaum, & Lilienthal, 1993) and were debriefed on the nature of the experiment upon completion.

The training phase of the experiment was a gamified novel word learning task. Participants learned the names of six novel objects. Names were selected from the NOUN Database (Horst & Hout 2016) and randomly assigned to the six objects. The objects were modelled to resemble potion bottles with visually distinct shapes and colors. The objects were arranged on either side of a large cauldron which occupied the center of the virtual space. Objects on the right side had handles on the right side and could only be grasped with the right hand, and symmetrically for the left side. Each novel object afforded either a left- or a right-handed grasp.

Participants were instructed to "Pour in these ingredients…" followed by one of the novel words. Following the prompt, participants picked up one of the objects by the handle (pulling the trigger on the controller to grasp) and tilted it over the virtual cauldron to pour the ingredient (**Fig. 2**). If the word matched the ingredient poured, a swirling particle effect indicated success. If the ingredient did not match, the cauldron exploded, and the potions were reset to the sides of the cauldron in random positions but without changing the side on which a given object appeared. Pouring trials were grouped into recipes of 2-6 non-repeating ingredients. If the participant correctly poured all of the ingredients in a recipe, a short musical tune was played, a virtual object (e.g. a floating globe) appeared somewhere in the environment, and the potions were randomly reset to their sides. The training process was repeated until participants completed 20 recipes. Most participants completed the training in 10-25 minutes. Two participants failed to complete the training phase in one hour and were excluded from further analysis.

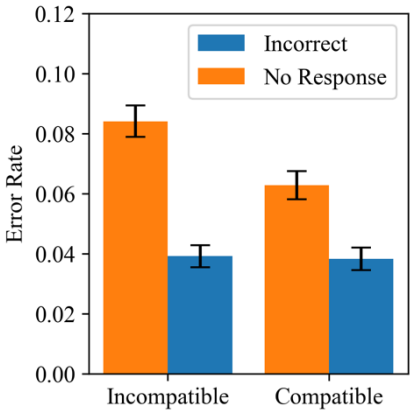
Fig. 2: Screenshots from the perspective of a participant in Experiment 1. (**Top**) Six novel objects arranged on either side of a cauldron. During each training trial, the participant heard one of six object names, then grasped one object by the handle (silver or gold rings) and poured it into the cauldron. (**Bottom Left**) If the correct object was poured into the cauldron, swirling "magical" particles would indicate success. A series of successful pours would cause an object (e.g. floating globe) to appear. (**Bottom Center**) An incorrect pour resulted in the cauldron exploding. (**Bottom Right**) During the test phase, the participant heard one of the novel words and saw a patch of color. The participant was instructed to pull the left (or right) trigger on the motion controller if the patch of color matched the color of the named object.

After training, participants performed a match-mismatch reaction time task based on the Action-Sentence Compatibility Effect (Glenberg & Kaschak 2002). Participants heard one of the words from the previous phase and were presented with a patch of color matching one of the objects. The patch of color was shown 100 ms after the start of the audio. Participants were instructed to respond as quickly as possible by pulling one controller trigger if the color matched the named ingredient or pulling the other if it did not match (counterbalanced between subjects). Trigger pulls performed during the match response were the same movements used to grasp the objects in the learning phase. If the word referred to a potion poured with the same hand as the trigger response, the trial was

coded as *compatible*, otherwise it was *incompatible*. There were 200 randomized trials, half of which were compatible and half of which were incompatible trials. There were also an equal number of *match* ("yes" response) and *mismatch* ("no" response) trials. If no response was made within 1.5 seconds, the trial ended and was recorded as a non-response. Trials were completed in 5 blocks of 40 trials separated by 10 second breaks. Word-object mappings were randomized in the training phase after every 10 subjects to control effects of word or color during the test phase. We recorded and analyzed which response was made and reaction times for all trials.

2.2. Results

A total of 5400 test trials were completed with a no-response rate of 7.4% and an incorrect response rate of 3.9%. We excluded mismatch trials from further analyses of affordance compatibility. Mismatch trials require a participants to retrieve both the object that the word references and the object that the color references. Either, none, or both of these objects may correspond to a compatible affordance. Match trials only require recalling the object referred to by both. We also eliminated incorrect and no-response trials. Response times that were 2.5 standard deviations away from the mean for each subject were discarded as outliers (less than 1% of the data). There was a small decrease in reaction times for most subjects during the first 10-20 trials indicating an effect of practice. However, we determined that practice effects did not interfere with further analysis, so we did not exclude initial trials.



Incorrect response rates were nearly identical for compatible and incompatible trials while the no-response rate was slightly greater for incompatible trials (**Fig. 3**). Both kinds of errors were relatively infrequent, which may be due to a lack of pressure to respond quickly. Compatible match trials were 29 ms (90% CI = (-5, 69) ms) faster on average for the right hand and 16 ms (90% CI = (-14, 44) ms) faster for the left hand.

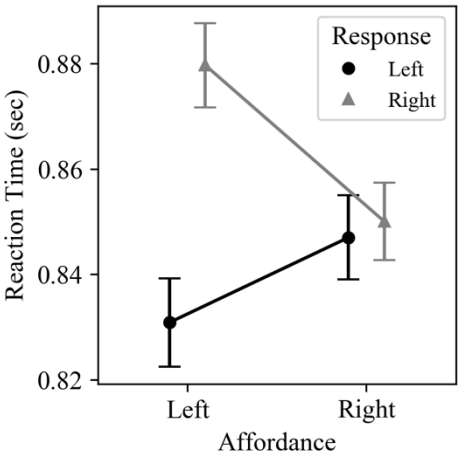


Fig. 3: (**Left**) The incorrect and no response rates in Experiment 1, for compatible versus incompatible trials. Error bars denote SEM. The no-response rate was roughly 2% greater for incompatible than compatible trials, corresponding to an average of 4 time-outs per subject. (**Right**) The mean reaction times for left and right responses by the affordance of the named object for match trials, excluding incorrect and no-response trials. Error bars denote SEM. The affordance compatibility effect is shown as an interaction between affordance and response.

We performed a linear mixed effects analysis on reaction times using R and the nlme package (Pinheiro et al., 2017), following recommendations from Zuur et al. (2009). Fixed effects included in the model were response (left or right), affordance (left or right), and the interaction between these variables. We added random intercepts for subjects, as maximum likelihood tests performed using REML (restricted maximum likelihood estimation) indicated this was the best fit for the random term of the model. We ran a likelihood-ratio chi-square test comparing the full model to a null model without the affordance by response interaction to determine whether the full model performed significantly better. We obtained p-values for individual model predictors by running the full model with REML. We then used a monte carlo simulation method to compute the probability of finding a significant interaction (ß = .88) and 90% confidence intervals for the model coefficients. We calculated the monte carlo estimates by sampling subjects and trials with replacement from our original data.

The full model significantly outperformed the null model ($p < .0001$, L ratio = 17.0). We did not find a significant main effect of response (coef = 54 ms, 90% CI [-12, 127], $t = 1.36$, $p = .18$) or affordance (coef = 19 ms, 90% CI [-10, 44], $t = 1.75$, $p = .08$). As predicted, we observed an interaction between response and affordance (**Fig. 3**), such that participants responding to matches with their left hand had quicker responses to words associated with the left affordance and participants responding to matches with their right hand had quicker responses to words associated with the right affordance (coef = -59 ms, 90% CI [-102, -11], $t = 4.13$, $p < .0001$, Cohen's d = .29). This interaction indicates that the novel words acquired an affordance compatibility effect: that actions which were compatible with the sensorimotor interactions practiced while learning a word were facilitated over incompatible actions.

3. Experiment 2: Action-observation induces affordance compatibility

We next asked whether the affordance compatibility effect for learned words could be induced without direct manipulation of objects. We conducted a second experiment in which participants performed the training task verbally while observing virtual hands manipulating the objects. Observation of actions recruits a network of brain

regions significantly overlapping with the areas active during execution of the same action (Hari et al., 1998; Buccino et al., 2001; Fadiga, Craighero, & Olivier, 2005; Grezes & Decety, 2001). Furthermore, observation of motor learning is found to facilitate motor learning in the observer upon later learning of the same task (Mattar and Gribble, 2005), suggesting that the observer was simulating the motor experience of the actor as they watched the action unfold. Thus, we expect the neural processes occurring in an action observation version of our task to substantially overlap with those in Experiment 1, resulting in similar formation of the associations between motor affordances and object labels. This would reduce the need for an individual to have exhaustive experience with a referent object or action in order to acquire fully grounded representations, since many grounded features can be acquired through social learning or observation of others.

3.1. Methods

Experiment 2 was conducted using the method from Experiment 1 with several modifications. In the training phase, following the prompt, instead of grasping and pouring one of the objects with the controller, participants verbally indicated which object they wished to pour. Each position from left to right was marked by a floating number (1 through 6). Participants indicated their choice by reading the number above the object. Participants were not given controllers. When the participant made a selection, the experimenter entered the choice on a keyboard and a virtual hand followed a pre-recorded trajectory to reach out, pick up, and pour the potion. Because the no-response rate was relatively high in Experiment 1, we increased the trial duration in the test phase from 1.5 s to 2 s to avoid truncating the reaction time distribution. We decreased the number of test trials from 200 to 160 due to concerns of fatigue, although our post-exposure comfort survey ultimately determined this was not an issue. Twenty-seven participants (18 women; 26 right handed; 18 bilingual; age 18-21 years) took part in this experiment.

3.2. Results

A total of 4200 test trials were completed with a no-response rate of 5.5% and an incorrect response rate of 5.2%. The mean reaction time for correct match trials was 981 ± 292 ms. As in Experiment 1, outliers were discarded. Likely due to the increased time to respond, no-response rates were lower in Experiment 2, and overall reaction times were greater and more variable. No differences were observed for no-response or incorrect response trials as a function of compatibility.

Statistical modeling was conducted as in Experiment 1. The full model outperformed the model without the interaction ($p = .036$, L-ratio = 4.39). We did not find a significant main effect for either response (coef = 31 ms, 90% CI [-78, 141], $t = -1.11$, $p = .62$) or affordance (coef = -17 ms, 90% CI [-56, 19], $t = 0.51$, $p = .27$). There was again a significant interaction between response and affordance (coef = 50 ms, 90% CI [-111, 21], $t = -2.09$, $p = .036$, Cohen's $d = .17$) (**Fig. 4**), consistent with our prediction that novel words can become associated with their affordances through action observation without action execution. After completing the study we calculated a power estimate using monte carlo simulation as in experiment 1 (ß = .538) suggesting that this

study design was slightly underpowered and a larger replication would help to determine the reliability of this finding.

The affordance compatibility effect in this experiment was driven only by right hand responses. Given that most of the participants were right-handed (96%), it is possible they were more likely to mentally simulate the reach and grasp movement performed with the right artificial hand. Furthermore, the effect size in this experiment was smaller than that of experiment 1. This may indicate that action observation alone gives rise to a weaker association between words and affordances.

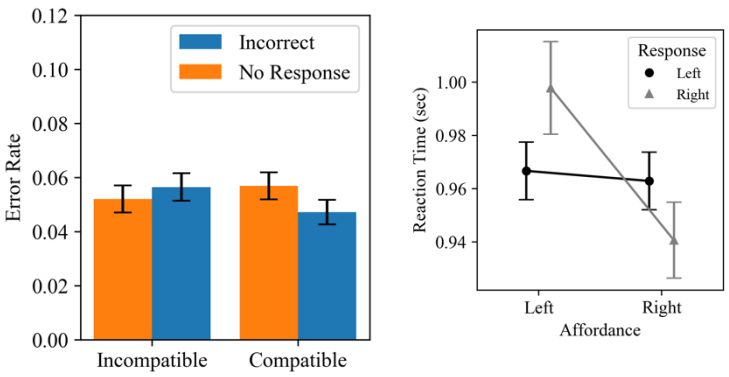


Fig. 4: (**Left**) The error rates for all participants in Experiment 2. There were no clear differences in error rates between conditions. With the increased trial duration, the overall rate of no-response trials decreased roughly 2% compared with Experiment 1, while the rate of incorrect responses increased 1%. (**Right**) The mean reaction times for left and right responses by the affordance of the named object in match trials. The interaction between affordance and response demonstrates an affordance compatibility effect, though largely driven by participants responding with their right hands.

4. Experiment 3: Space and hand interact in affordance compatibility

In Experiments 1 and 2, the affordance of each novel object was represented redundantly through the position of the object and the orientation of the handle, as well as through corrective instructions to participants if they attempted to use the incorrect hand. This is consistent with many natural interactions with objects, in which both spatial and visual features indicate affordances. However, because the relative location of an object and the hand used to interact with it was always consistent, it was not possible to distinguish between the contribution of the specific hand and the side of space in which the interaction occurred. The affordance compatibility effects we observed in Experiments 1 and 2 could have been caused by either factor. Therefore, in a third experiment, we separated the spatial and hand compatibility dimensions by swapping the positions of some objects during training. We suspected that the affordance compatibility effect in Experiments 1 and 2 was primarily driven by handedness and we would observe a significant interaction between affordance hand and response hand, but we did not have any *a priori* hypotheses regarding spatial compatibility.

4.1. Methods

The methods were similar to those of Experiments 1 and 2, aside from the location of each of the novel objects during training. One left-handed object always appeared on the left side and could only be picked up with the left hand. Another always appeared on the right side and could only be picked up by the left hand, requiring

participants to reach across their body. A third object was always picked up by the left hand, but randomly alternated between the left and right sides. Right-hand objects had a corresponding flipped arrangement. During the test phase, the hand used to respond was either compatible or incompatible with the hand used to grasp the named object (hand compatibility) and was either on the same side of space, the opposite side of space, or mixed (space compatibility). To ensure the participants knew which hand to use for each object, the object handles were adjusted to face prominently in the direction of the correct hand, and the participants were instructed to pick the object up with the hand matching the handle direction. To collect sufficient data for the factorial design (2 hand x 3 space), 43 participants (34 women; 39 right handed; 26 bilingual; age 18-41 years) completed the training phase and 200 test trials each.

4.2. Results

Participants in Experiment 3 completed 8160 test trials with an overall no-response rate of 3.8% and incorrect response rate of 7.3% (no significant differences by trial types). The mean reaction time for correct match trials was 927 ± 291 ms. Test trials in which the response hand was compatible with the hand used to grasp the named object were 22 ms faster than incompatible trials. Spatially incompatible trials were 9 ms faster than mixed trials and 13 ms faster than spatially compatible trials (**Fig. 5**).

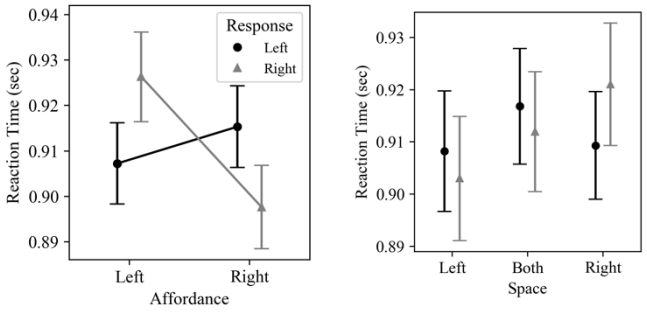


Fig. 5. (**Left**) Mean reaction times (correct match trials) for left and right responses by the hand used to grasp the named object. The hand compatibility effect is similar to the overall affordance compatibility effect in Experiments 1 and 2. (**Right**) We did not find any significant interaction between response hand and spatial affordance. Spatial compatibility did significantly interact with hand compatibility.

As in Experiments 1 and 2, we applied an iterative model testing procedure to choose the best linear mixed effects model determined by model fit and complexity but including the additional interactions with spatial affordance. The model fit procedure and related data processing can be found in the code included in supplementary materials. The random structure of the model was the same as that of Experiments 1 and 2. Fixed effects included in the optimal model were response hand, spatial affordance (left, right, both), hand affordance (left, right), the interaction between space and hand, and the interaction between hand and response. Our model testing indicated that including the interaction between space and response did not significantly improve model fit ($p = 0.13$, L-ratio = 4.06) therefore this interaction was omitted from the optimal model. The final model significantly outperformed a model without the hand by response interaction ($p < 0.0001$, L-ratio = 11.3, ß = .78) and without the hand by space interaction ($p < 0.0001$, L-ratio = 14.9) (**Table 1**). There was no significant main effect for response or hand

affordance. We did observe a main effect of left spatial affordance. As in Experiments 1 and 2, we observed an affordance compatibility effect for the affordance hand (Cohen's d = -0.19). We also observed an interaction between hand affordance and space affordance, where having the hand and space share the sidedness feature (both right or both left) speeds responding to the corresponding word (Cohen's d = 0.24). These results are consistent with our prediction that hand compatibility is likely the more significant factor in the effects observed in Experiments 1 and 2. Although we did not make predictions regarding the effects of space on reaction time, our results indicate that space might modulate the hand-specific effects. We discuss potential explanations for this below.

Table 1

Linear mixed-effects model fit by REML

	Value (s)	Std. Error	DF	t-value	p-value	90% CI
Intercept	1.050	0.038	3457	28.0	0.0000	[.997, 1.139]
Right Response	-0.023	0.051	41	0.45	0.65	[-.088, .025]
Right Hand	-0.021	0.017	3457	-1.25	0.21	[-.079, .099]
Left Space	-0.040	0.014	3457	-2.80	0.005	[-.101, .019]
Right Space	-0.012	0.014	3457	-0.81	0.42	[-.057, .032]
Right Response * Right Hand	-0.056	0.017	3457	-3.36	0.0008	[-.109, -.003]
Right Hand * Left Space	0.075	0.020	3457	3.71	0.0002	[.015, .185]
Right Hand * Right Space	0.019	0.020	3457	0.94	0.35	[-.040, .095]

5. Discussion

We conducted three novel word learning experiments using virtual reality to investigate the ways in which object affordances become associated with words. Participants learned the names of virtual potions through interaction in several training conditions. Then, in a word-color matching task, we observed faster responses for words which referred to objects grasped with the same hand used for the response than those grasped with the opposite hand. We refer to this as an affordance compatibility effect.

This is an important demonstration of naturalistic sensorimotor interaction during word learning giving rise to embodied language effects.

In our second experiment, we confirmed that an affordance compatibility effect can be induced, perhaps to a lesser extent, through observation of virtual effectors. This result is consistent with theories suggesting that action observation networks in the brain support imitation and social learning (Iacoboni, Woods, Brass, Bekkering, Mazziotta, & Rizzolatti, 1999), and suggests that words and concepts learned while watching others may activate sensorimotor experience-dependent networks in the same way as words learned by doing. Prior work has shown that motor learning can occur during passive observation of a motor task (Mattar & Gribble, 2005), but these findings further demonstrate that action observation influences language learning. It is important to note again, however, that this observation network does seem to activate this embodied representation to a weaker degree than direct manipulation. While the affordance compatibility effect was observed for the right hand responses, it was absent for left hand responses. Research shows that right-handers may have a more difficult time learning words that represent left-handed actions (Nooijer et al., 2013), which may play a role in our study. Additional work will be needed to explore the role and limits of observation as a means of acquiring embodied semantic knowledge.

In our third experiment, we sought to understand the relationship between effects of hand versus space on the affordance compatibility effect. We confirmed that the effector used to interact with an object acquired an affordance compatibility effect even when crossing the body to interact. We did not identify a direct relationship between the space in which affordances were learned and the resulting response dynamics. The interaction between space and hand suggests that a direct study of the relationship between spatial and motor affordances would be beneficial. Response times were faster when spatial and effector-specific affordances were consistent, regardless of the hand used to make the response. This suggests that a shared affordance feature in this context might facilitate response preparation. Given that spatial location frequently corresponds with effector-specific affordances, interacting with and learning about objects for which these features are inconsistent may engage distinct cognitive and neural mechanisms. The present study was not able to test these predictions directly, preventing a clear picture of spatial affordance effects from emerging. More work is needed to address these questions. This experimental paradigm introduces a platform which can be extended for further exploration of cognition grounded in naturalistic body movements.

Another explanation for these results that we considered is that the affordance compatibility effect is related to a Simon effect (Simon, 1969; Roest et al., 2016). The Simon effect refers to a pattern of faster responses when the response and the stimuli share an overlapping spatial dimension. This widely replicated finding can be seen, for instance, by showing participants a picture of a mug, where the participant needs to push a button with the left hand to classify the object. In this example, the participant should be faster to respond if the handle of the mug faces left than if it faces right. In our experiments, an overlap between the response and the handle direction of the recalled object resembles a Simon task. Beyond this superficial similarity, however, our paradigm bears little relationship to the Simon task. The Simon effect has only been reported when the visual stimulus bearing a spatial indicator co-occurs with the response. In our

experiment, participants recall the novel object from memory given only a verbal cue. Further, in the Simon task, the spatial cue of the stimulus is carefully presented within the visual field, whereas in our tasks participants are free to look around naturally and very often do not look at the objects as they hear the words. We therefore do not believe these results can be explained by the Simon effect, nor do we think it is conceptually straightforward to extend the Simon effect to encompass these findings. Nevertheless, it is important to situate our findings among the broader literature on action compatibility.

One important caveat of this work, and much of the existing research on embodied language, is that it is difficult to determine precisely how interactions between language and sensorimotor systems give rise to action compatibility effects. These effects are typically small perturbations of response latency or accuracy which could be caused because cognitive conflict is induced by incompatible action representations or because motor preparation is facilitated by linguistic activation of motor regions. In either case, response incompatibility is obviously regularly overcome during natural behavior, so it is difficult to know the importance of affordance compatibility for everyday cognition. These challenges are not unique to embodied language research: visuospatial compatibility effects (e.g. S-R Compatibility, Michaels, 1988) similarly rely on slightly speeded responses to investigate motor representations. Nevertheless, a more direct test of the efficacy of sensorimotor activity in linguistic representations is needed if embodied language accounts are to replace, rather than complement, amodal symbolic representations.

A crucial piece in the understanding of embodied language will come from bridging short-term embodied learning effects like those demonstrated here with longer-term embodied language effects observed in fluent adults (Glenberg & Kaschak, 2002; Barsalou, 1999). Embodied language theories predict a progression from specific sensorimotor associations to more flexible, generalized sensorimotor simulations as words are expressed in a broader set of contexts (Zwaan, 2004; Barsalou, 2009). Many unanswered questions remain about how associations between affordances and words change over weeks and months of sensorimotor experience. The effort to answer these questions will benefit from the integration of naturalistic infant and child language learning research with virtual reality experiments offering greater control over sensorimotor interactions. As these questions are tackled, we may come to better understand how our bodies and environments give meaning to the words we use.

Chapter 5
Multimodal Music Perception Engages Motor Prediction: A TMS Study.

Corticospinal excitability (CSE) in humans measured with Transcranial Magnetic Stimulation (TMS) is generally increased by the perception of other people's actions. This perception can be unimodal (visual or auditory) or multimodal (visual and auditory). The increase in TMS-measured CSE is typically prominent for muscles involved in the perceived action (muscle specificity). There are two main classes of accounts for this phenomenon. One suggests that the motor system mirrors the actions that the observer perceives (the resonance account). The other suggests that the motor system predicts the actions that the observer perceives (the predictive account). To test these accounts (which need not be mutually exclusive), subjects were presented with four versions of three-note piano sequences: sound only, sight only, audiovisual, and audiovisual with sound lagging behind (the prediction violation condition). CSE was measured in two hand muscles used to play the notes. CSE increased reliably in one muscle only for the prediction violation condition, in line with the predictive account, while the other muscle demonstrated CSE increase for all conditions, in line with the resonance account. This finding supports both predictive coding accounts as well as resonance accounts of motor facilitation during action perception.

Introduction

Motor regions of the brain are traditionally defined by their primary role in motor control (i.e., coding goals, planning, coordinating, and executing actions) but motor areas additionally play a role in the perception of others' actions (e.g., Hari et al., 1998; Buccino et al., 2001; Aziz-Zadeh et al., 2004; Fadiga et al., 2005). A common measurement used for detecting motor activation is transcranial magnetic stimulation (TMS) -induced motor-evoked potentials (MEPs), which reflects the level of corticospinal excitability (CSE) at the time of stimulation. MEPs have high temporal resolution, allowing for a precise measure of activity modulation. Increased CSE is found during visual perception of actions (Fadiga et al., 1995) as well as auditory perception of actions (Kohler et al., 2002; Aziz-Zadeh et al., 2004). This increase is thought to reflect the recruitment of the mirror neuron system (Gallese et al., 1996), which is active both during action observation and action execution for similar actions, suggesting its involvement in the understanding of others' actions (e.g., Fogassi et al., 2005; Iacoboni et al., 2005; Kaplan and Iacoboni, 2007; Gallese, 2008; Kilner, 2011). Populations of mirror neurons have been uncovered in premotor cortex that discharge upon observation and execution of the same action (Rizzolatti et al., 2001; Ferrari et al., 2003). The increased excitability in action perception is additionally time-dependent and effector-specific relative to the action being observed. Gangitano et al. (2001) recorded MEPs from the first dorsal interosseous (FDI) muscle during observation of a cyclic hand movement, and found that at the time when the observed finger aperture was at its maximum, MEPs recorded from the observer's FDI muscle were highest. When the observed finger aperture was at its minimum, MEPs were lowest. Thus, the cortical motor areas of an

observer are recruited for motor simulation of others' actions in synchrony with those actions, and this is specific to the same muscle involved in the action.

Motor activity during action observation is also referred to as motor resonance, due to its time-dependent and effector-specific nature. The motor system of the observer "resonates" with that of the actor, allowing the observer to use their own body to understand, from within, the action being performed. One unanswered question is whether action observation is an active and predictive top-down process or a more automatic, bottom-up process. Early accounts, such as the direct-matching hypothesis (Iacoboni et al., 1999), suggest that motor resonance arises by directly perceiving an action via automatic activation in the observer of the cortical areas that represent the execution of that action. More recent theories propose that the mirror neuron system may function as a prediction mechanism during observation of others' actions (Kilner et al., 2007). This differs significantly from the traditional assumptions of simulation (Gallese and Goldman, 1998), where the body of the observer "resonates" with the observed action but does not actively predict the future states of the movements. Active prediction proposes that top-down mechanisms influence the increase in CSE during observation, as a way to follow along with and actively predict another's movements. It may be that mirror system activity actually reflects this predictive process, as the brain uses what it knows about the motor system of the observed actor to project the future state of the actor's body. Because this type of prediction is very similar to that in motor control, the same neural systems (i.e., the motor system, mirror system) will underlie this process. Essentially, an observer can predict the motor commands of an observed actor given the expectations about their goal, and the implemented kinematics of that movement can be predicted using the observer's own motor system.

Increased activity in motor areas has been observed in pianists when listening to piano pieces and observing piano playing (e.g., Haueisen and Knösche, 2001; Bangert and Altenmüller, 2003; Meister et al., 2004; Haslinger et al., 2005; Bangert et al., 2006; D'Ausilio et al., 2006), suggesting involvement of the mirror neuron system during music perception. Researchers have thus proposed that music is not passively heard, but actively perceived as the expressive motor acts that caused the music and are instantiated in the mirror neuron system (Molnar-Szakacs and Overy, 2006; Wallmark et al., 2018). Music-making observation is also a good candidate domain for exploring prediction during action observation, due to its sequential, and thus predictive, nature. Candidi et al. (2012) found that when expert pianists observed a fingering error (a note played with the incorrect finger), CSE recorded from the muscle corresponding to the finger playing the note increased significantly compared to the correct fingering of the keys. Non-musicians who were visually trained to detect the errors did not show this muscle-specific increase in CSE during the fingering errors. The authors conclude that the experience of musically trained pianists provides their brains with *simulative error monitoring systems*. In other words, when participants were observing the fingering error, a prediction error occurred, leading to an increase in motor system activation. Furthermore, Stephan et al. (2018) showed in a recent study that auditory cues from a learned melody led to increased activation of the muscle that plays the *following* note of the melody, suggesting an anticipatory process occurring during melody processing in music perception.

Unexpected fingering errors cause one kind of prediction error that reflects an error of intent: the observer assumes that the player will play with one finger, and this prediction is violated when the player uses another. There are also purely sensory errors that can give rise to prediction errors, such as a multimodal stimulus with a misalignment between the auditory and visual components. Sensory errors that are not tied to specific effector movement error may not use the motor prediction system, as the prediction might differ from motor control prediction. On the other hand, existing research (Schubotz, 2007) has found that prediction of non-human-created sensory states (i.e., pitch prediction, object prediction) also relies on the motor regions of the brain. This suggests that we may see increased predictive activity in motor areas during a sensory prediction error as well.

We aimed in our study to explore CSE modulation when there is no error in human movement production, but the sensory consequences of observed movements are temporally misaligned, resulting in a sensory prediction error. Specifically, the auditory correlate of a piano key press is delayed so that it begins 550 ms after the visual key press occurs. As the incoming visual signal of the motor act is perceived, a prediction of the corresponding auditory consequence is made. When the auditory signal is delayed so that onset is 550 ms into the video, the sensory prediction is violated. If the motor system is involved in sensory prediction, this stimulus should increase CSE due to error detection. Systems that use predictive coding mechanisms work optimally by using low resources when predictions are closer to the actual observed state, and increasing resource use during large discrepancies or errors (Mlynarski and Hermundstad, 2018). If effector-specific motor regions are recruited for generating sensory predictions, we can expect to observe an increase of CSE in the observed muscle during sensory discrepancies, but not an increase in other muscles (effector-specificity). For example, if a participant is observing a key played with the actor's index finger, we should see increased excitability recorded from the participant's index finger muscle, but not from their pinky muscle. Furthermore, while detection of fingering errors increases CSE in experienced musicians only, the prediction error in our study should give rise to facilitation in both musicians and non-musicians, as no training is needed to understand the relationship between the observed action and the timing of its sensory consequences (c.f. Candidi et al., 2012).

In addition, while it is known that both visual and auditory action observation leads to increased CSE, the differential influences of each of these modalities on their own remains unclear. It is also unclear whether multimodal action perception will lead to additive activity in motor cortex summing over both visual and auditory contributions to motor regions. If motor involvement is primarily a bottom-up, automatic phenomenon, we might expect additive effects of multimodal presentation on CSE. The active inference framework, however, predicts that multimodal presentation will not lead to increased CSE over single-modality, as the information is redundant for predictive purposes. Thus, if we see summative effects, we have evidence for the resonance account. On the other hand, if we do not see additive properties, we cannot rule out either theory as there may also be ceiling effects or over-dominance by one of the sensory modalities. Therefore, an additional objective of the present work was to explore the modulation of CSE during auditory, visual, and multimodal music perception to explore the potentially different effects of modalities on motor system activity.

Methods

Participants

Forty subjects were recruited for this study (15 males, mean age = 22 ± 2.6, range = 18–26). Due to excessive noise in the EMG signal or a participant having difficulty maintaining wakefulness, we excluded two subjects. Twenty-three of the final participants were non-musicians and had never played the piano, five had played for less than 3 years and 10 for more than 3 years. All subjects were right-handed and had normal or corrected-to-normal vision and no hearing impairments. Subjects were screened for contraindications for TMS and previous medical problems that would be risk factors for TMS. The UC Merced Institutional Review Board approved the study, and written consent was obtained for all subjects. The experiment took about 1 h, and subjects received two research credits that can be used for credit in some undergraduate courses.

Stimuli

Stimuli consisted of video recordings of a three-note piano sequence played using the right hand. Half of the stimuli were recordings of the thumb finger playing one note, followed by the index finger playing the next note twice. The finger movement that produces this action involves the FDI muscle. The other half had the same pattern but played by the ring finger followed by the fifth digit (pinky). This second half was played one scale higher than the keys used in the first half, so that the auditory cue for each could be distinguished. The finger movement that produces this action involves the flexor digiti minimi (FDM) muscle. For each of these two patterns, we created auditory-only (video blacked out), visual-only (audio silenced), multimodal, and multimodal time-lag versions, using iMovie. The time-lag version was created by starting the audio for the video 550 ms after the visual stimulus began. Eight different stimuli videos were used, resulting in a 2 (finger movement: thumb-index-index, ring-pinky-pinky) by 4 [modality: Auditory (A), Visual (V), Audiovisual (AV), AV-lag] design. The videos were filmed using an iPhone 7 Plus camera (resolution 1080p; 60 frames per second) and were edited using iMovie. Each video was played 10 times in a randomized order, leading to a total 80 stimulations during the experiment. TMS stimulation was triggered by our presentation software, Paradigm, at the time in which the index or pinky finger began its second down press, 1400 ms into the video. In between trials, a crosshair appeared on the screen where participants were instructed to maintain focus. **Figure 1** displays a visual depiction of the trial sequence.

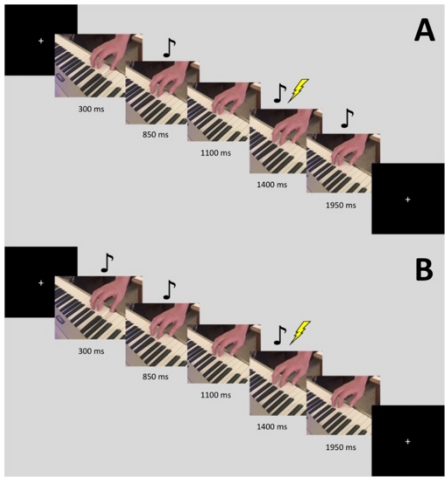


Figure 1. Sequence for a single trial with the FDI stimulus. The images are slices taken from the video. Stimulation occurred 1400 ms into the video. The music notes denote when the auditory tone was heard. Panel **(A)** displays the sequence for the AV-lag condition, while panel **(B)** displays the sequence for the regular AV condition.

TMS and EMG Recording

Corticospinal excitability was measured by the peak-to-peak amplitude of motor evoked potentials (MEPs) recorded using electromyography (EMG) on two muscles of the right hand. Two bipolar surface electrodes were placed on the belly of the participants' right FDI muscle. Two additional electrodes were placed over the FDM muscle. A ground electrode was placed on a bone near the elbow of the subject. In order to obtain optimal EMG signal, we abraded and cleaned the skin under the electrodes, and secured the electrodes with medical tape. A bandpass filter (50–1000 Hz) was applied to the EMG signal, which was digitized at 1024 Hz for offline analysis. MEPs were elicited by applying single-pulse TMS to the region of the left motor cortex that induced MEPs in both FDI and FDM. If a location that induced MEPs in both muscles could not be determined, we used the FDI hotspot and thus did not record MEPs from the FDM muscle. Pulses were delivered using a Magstim Rapid2 TM with an attached 70 mm figure-of-eight coil positioned over the optimal scalp location with the handle pointing backward at 45° from the midline. The motor hotspot localization procedure was as follows. Subjects were fitted with a swim cap that was covered by a grid of dots 1 cm apart. Optimal scalp position was determined by moving the coil in 1 cm intervals until the location eliciting the best MEPs in both muscles was identified. We were unable to find the shared hotspot position for six subjects, and thus only have data from FDI for these subjects. The optimal location was marked on the swim cap worn by the participant. Resting motor threshold was determined as the percent of machine output that produced 3 out of 6 MEPs of at least 50 mV peak-to-peak amplitude. The stimulation intensity during the experiment was set to 120% of a participant's resting motor threshold. The coil was held steady at the optimal position throughout the experiment. The inter-pulse interval between each stimulation was between 9 and 10 s. Subjects were instructed to keep their head still and remain relaxed with their right hand on their lap for the duration of the experiment, while attending to the videos as they appeared.

Results

The EMG data was exported from Visor2 (ANT Neuro), and we ran a custom Python script to extract MEPs (peak-to-peak amplitudes). We also calculated area under the curve, but as these values correlated over 98% with the peak-to-peak amplitudes, we did not use both measures. In order to use inter-individual comparisons, *z*-scores were calculated separately for each muscle from each participant. Trials in which MEP

amplitudes were larger than 2.5 standard deviations from the mean and those less than 50 μV were excluded as outliers. Less than 5% of all data were excluded. Statistical analyses were carried out in R.

A repeated-measures analysis of variance (ANOVA) was conducted on MEPs from each muscle to assess the significance of the effect of our experimental conditions on the MEP amplitudes. We had a repeated-measures 2 × 4 design with two finger movements (FDI, FDM) and four modality conditions (A, V, AV, AV-lag).

The amplitude of MEPs recorded from FDI was significantly modulated by the modality of the observed action [main effect: $F(3,111) = 3.52$; $p = 0.01$] (**Figure 2A**). This effect is due to the AV-lag condition (the prediction violation condition) inducing significantly larger MEPs (z-score = 0.103 ± standard deviation: 0.055) than the other modalities (z-scores; A: -0.067 ± 0.052, V: -0.032 ± 0.05, AV: 0.016 ± 0.05). We also observed an interaction between finger movement and modality [$F(3,111) \pm 4.93$, $p <$ 0.01], meaning that we did see muscle-specificity in MEP modulation in some conditions but none or less in others. We observed no additional main effect of the finger movement condition. *Post hoc* multiple comparisons using Tukey's honest significant differences revealed that when observing the index finger pressing the key, the AV-lag condition (the prediction violation condition) produced larger MEPs in the FDI (z-score = 0.223 ± 0.055) than while observing the pinky finger pressing the key (-0.017), [$t(68) =$ 0.391, $p < 0.05$], while there were no significant differences in the other conditions.

The amplitude of MEPs recorded from FDM is significantly modulated by the observed finger playing the note, [$F(1,31) = 4.32$, $p < 0.05$] (**Figure 2B**). This muscle-specificity in MEP modulation is driven by larger MEPs when observing the pinky finger pressing the key (z-score = 0.40 ± 0.064) compared to while observing the index finger pressing the key (z-score = 0.33 ± 0.053). Modality was marginally significant [$F(3,93) =$ 2.22, $p = 0.08$], with the AV-lag condition resulting in larger MEPs (z-score = 0.076 ± 0.07) than the other conditions (z-scores; A: 0.008 ± 0.065, V: -0.083 ± 0.06, AV: -0.01 ± 0.06). We did not obtain an interaction between modality and finger movement.

Average normalized MEP amplitudes for each modality condition and finger movement can be seen in **Figure 2**. Overlaid example FDI MEPs from the two finger movement conditions (index and pinky) in the AV-lag modality are given in **Figure 3**.

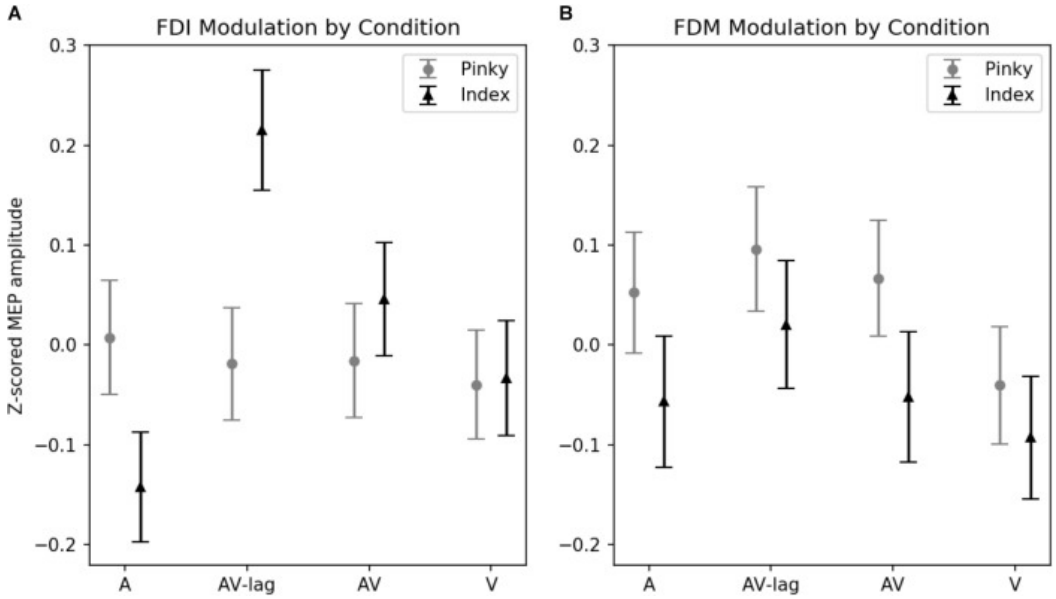


Figure 2. *Z*-scored MEP amplitudes for each modality and finger movement. Data from all subjects. Vertical bars denote standard error of means. **(A)** MEPs recorded from FDI. Motor evoked potentials in the index-press AV-lag condition show the largest facilitation. **(B)** MEPs recorded from FDM.

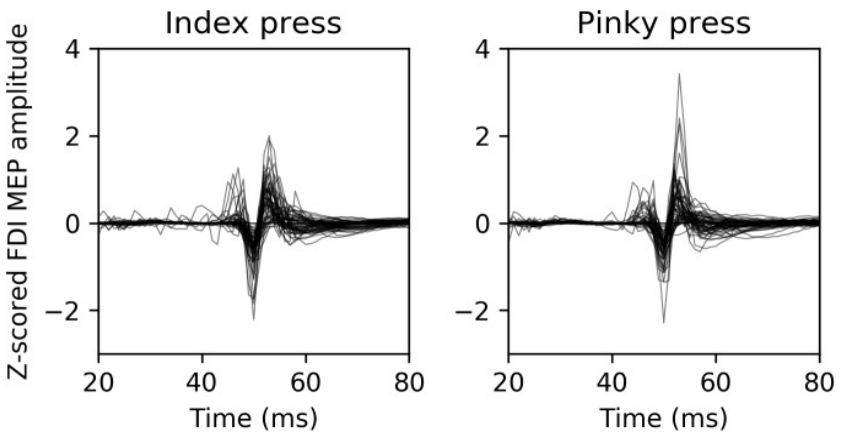


Figure 3. Representative example MEPs recorded from FDI from 13 randomly selected participants during the AV-lag trials. The left image contains MEPs from index press observation condition, and the right image contains MEPs from pinky press observation condition.

Discussions

In this study, we tested whether the increase in CSE during action perception is driven by bottom up resonance or by top down predictive coding. A selective increase in muscle specific CSE for the AV-lag (the prediction violation) condition that violate sensory expectations, would support the top down, predictive coding hypothesis. The FDI data support the predictive account, by showing an interaction between finger movement and modality, with larger MEPs for the AV-lag index movement stimuli compared to the

AV-lag pinky movement stimuli. The FDM results, on the other hand, are consistent with the motor resonance account, as we observed muscle-specific modulation in all conditions, and no interaction between observed finger movement and condition, with and without sensory expectations violation.

One potential explanation for this difference is that motor cortex representations of muscles that are used more often and for actions that require more skillful movements may have a higher predictive role during action observation. The FDI muscle is a heavily used muscle for object-oriented, manipulative, skilled actions. On the other hand, the FDM is used in a significantly less specific way. Indeed most humans are unable to move their pinky finger without coactivating at least the ring finger. Further studies may test this hypothesis.

Corticospinal excitability increased the most in FDI for the AV-lag FDI stimulus, suggesting that the sensory error was detected and processed in motor regions. Prior work (Candidi et al., 2012) has shown that the motor system in musicians enables simulation of observed piano playing, and that activation increases when a fingering error by the pianist is observed. Here we provide evidence that motor activity also increases during the observation of a non-movement-related sensory error during action observation, which is present even in our participant pool, comprised primarily of non-musician subjects. The delayed onset of the auditory component resulted in a sensory prediction error, and a corresponding increase in CSE. This provides evidence for a general predictive process taking place in motor areas of the brain at multiple levels, from intention prediction to sensory consequence prediction.

Active inference or "action-oriented predictive processing" has gained much interest over the last few years as a potential framework of how the brain instantiates perception, action, and cognition (Miall, 2003; Friston et al., 2011; Clark, 2015). Instead of considering the brain as a passive processor of bottom-up sensory information, these theories suggest that the brain is undergoing top-down active inference in order to predict incoming sensory information. Sensory information that is received provides feedback for top-down predictions to adjust predictive models in order to decrease prediction error in the future. Under this framework, prediction happens at multiple levels. At each level, generative models are created to predict information about the upcoming state of a lower level. Generative models calculate a prediction error based on a comparison of expected to actual sensory state. The prediction error is sent up the hierarchy, so that top-down mechanisms can adjust future predictions. This recurs until prediction error of the system is minimized. An important conceptual distinction in active inference theories is that motor processing is no different from sensory processing, as both are involved in top-down processing/prediction. A resulting idea from this is the existence of a single action-perception process that attempts to predict sensory input from all modalities. For action, the modality being predicted is proprioceptive input. The primary goal is to minimize surprise and thus minimize prediction errors.

In active prediction, neurons that are typically known to represent particular actions also represent the causes of sensory input (the same idea underlies ideomotor theory; Prinz, 2005; Hommel, 2013). In other words, perception and action share a common neural code. As such, Friston et al. (2011) suggest that the mirror neuron system can also be explained with active inference and predictive coding. Active inference

implies a circular causality, whereby actions are deployed in order to fulfill predictions prescribed by perception, which updates these predictions using information obtained via actions. During action observation, the same process is instantiated, but without the corresponding proprioceptive feedback that occurs during action. This means that the same neuronal ensembles that encode an action during movement will encode that same action during observation. This naturally allows for the formation of mirror neurons, which will underlie this predictive process (Kilner et al., 2007) for both action observation and action execution. Neurons with mirroring properties have now been described in multiple systems of the primate brain. Beyond the original findings in fronto-parietal circuits for grasping (Rizzolatti and Craighero, 2004), mirroring response have been recorded in dorsal premotor and primary motor cortex for reaching movements (Dushanova and Donoghue, 2010), in the lateral intraparietal area LIP for eye gaze (Shepherd et al., 2009), in the ventral intraparietal area VIP for touch (Ishida et al., 2010), and in human SMA and medial temporal cortex for grasping actions and facial expressions (Mukamel et al., 2010). This pervasive mirroring machinery seems ideal for generating predictive models during action observation. Higher-level generative models will make predictions about intentions and goals, while levels lower in the hierarchy will be involved in prediction of observed low level muscle movements.

This kind of predictive mirror neuron system can explain our results as well as the increased motor activation during fingering errors reported in Candidi et al. (2012), where error detection caused increased activation in predictive models to account for perceived error. When the visual component of the AV-lag stimulus begins with no auditory counterpart, the prediction is that the given trial is a visual stimulus only trial. At the auditory component onset, this prediction is violated and there are misaligned sensory representations of an ongoing observed action. Future studies may manipulate experimentally the number of trials that violate expectations within experimental blocks to further test the predictive coding hypothesis, as previously done for action preparation (Bestmann et al., 2008).

In summary, sensory error detection during action observation leads to increased activity in FDI. This facilitation likely results from prediction error caused by a mismatch between expected sensory consequence and actual sensory input. Sensory prediction errors may be generated in motor regions, and potentially rely on mirror neurons for this predictive process. This suggests that reconsidering the mirror system not only as a passive simulation mechanism, but also as supporting predictive mechanisms, may help improve our understanding of the functions of this system. On the other hand, FDM modulation seemed to reflect motor resonance, as muscle-specificity was the only significant predicting variable. This is not surprising, since there is no reason to think that motor resonance and predictive coding cannot co-exist. Indeed, a previous study by Gangitano et al. (2004) reported results more in support of a pure motor resonance function of the mirror system, in which the “surprising” conditions violating expectations did not induce any modulation of MEPs.

Future studies may want to determine under which conditions the mirror system and/or motor areas employ active prediction or motor resonance. Our experiment used music as a tool to explore this question. It is possible that music is particularly special in its multimodal, sequential nature. It remains to be tested whether different kinds of

sensory prediction also involve active prediction in the motor system, or if other contexts (i.e., novel actions) may invoke a more passive, resonant role. Future work is warranted to investigate other kinds of sensory prediction and involvement of the motor system in these domains.

Chapter 6
Recruitment of the motor system in the perception of handwritten and typed characters.

Many different functional roles have been ascribed to the motor system due to its prevalent recruitment in perceptual and cognitive tasks other than motor production. We discuss findings that suggest the motor system might take on multiple roles that vary with context and the brain networks involved. Using single-pulse TMS, we measured the corticospinal excitability of the FDI muscle in primary motor cortex as participants viewed words that were either typed or handwritten. We observed consistent facilitation of corticospinal excitability during reading of handwritten text. Although we observed facilitation in corticospinal excitability during the presentation of typed text, this effect decreased with repetitive presentations of stimuli. We suggest that the facilitation during presentation of typed words is a case of action simulation, and that the diminishing facilitation in the case of typed stimuli is representative of sensory prediction by the motor system. These findings suggest that we should consider multiple roles for motor recruitment during the observation of visual stimuli, taking context into consideration.

Introduction

The motor system is involved in a large number of cognitive and perceptual domains, including action observation, perception of object affordances, speech perception, language and metaphor, and social cognition. There are many theories aimed at explaining this widespread use of the motor system. We will introduce some of these theories here and work supporting each of them. Then we provide an alternative hypothesis: that there is no one particular role for the motor system in perception and cognition, but that it plays many roles decided, in part, by situational context.

Outside of its role in moving the body, the most common and widely posited role of the motor system is in observation-execution mapping. A network which includes several motor regions of the brain is responsible for mapping observed actions onto one's own motor system, which is said to underlie action understanding. The neurological underpinnings of this system rely on particular neurons in motor cortex, called mirror neurons, that fire during both observation and performance of a motor act in macaque monkeys and in humans. (Rizzolatti et al. 1988, di Pellegrino et al., 1992; Gallese et al., 1996; Mukamel et al., 2010). Umiltà et al. (2001) found that some subset of mirror neurons fire during the final part of an observed action, even if that final part of the action is occluded from view, suggesting that mirror neurons code the goal-related execution of an action. This also suggests that mirror neurons respond to action-related situations where determining the actor and situation involves more inference, suggesting a role in deeper understanding of action. Kohler and colleagues (2002), recording from single neurons in monkey premotor cortex, found that some of the same neurons that fire during observed action will also fire when monkeys are only hearing the auditory information from the action (i.e., the cracking of a peanut). Again, this involves inference to the presence of the actor without visual recognition.

Motor activation during action observation is also called motor resonance (Iacoboni, 1999), due to its timedependent and effector-specific nature. It is said that the

motor system of the observer "resonates" with that of the actor, allowing the observer to use their own body to understand the action being performed. Gangitano, Mottaghy, and Pascual-Leone (2001) applied transcranial magnetic stimulation (TMS) to record motor-evoked potentials (MEPs) from the first dorsal interosseus (FDI) muscle on the right hand during the observation of a cyclic hand movement. They found that at the time when the FDI muscle of the observed hand was most contracted, MEPs in the observer were highest, and when the muscle was least contracted, MEPs were lowest. Thus, the motor resonance occurring in the observer is timelocked with specific muscle activity in the observed agent.

A related theory of motor system involvement is that of overt action simulation (Barsalou, 2009; Gallese and Lakoff, 2005). This is related to the above mentioned position and not mutually exclusive, as observation execution matching could involve low-level simulation of an actor. Simulation theories, however, posit explicit ongoing simulation underlying perceptual and cognitive processes as a sort of online enactment, particularly for understanding semantics of action language. In other words, linguistic phrases such as "the boy caught the ball" are understood by low level simulation of the action in the sentence. Numerous studies have shown that action language activates the motor system. In an fMRI experiment, Hauk, Johnsrude, and Pulvermuller (2004) found that when participants read action words related to the arm, leg, or face, the corresponding regions of the motor somatotopy were active. Oliveri et al. (2004), using single pulse TMS, found that motor cortex activation increased for both action nouns and action verbs when compared to activation during non-action words. Candidi and colleagues (2010) found that verbs conjugated in the future tense induce higher corticospinal excitability than verbs conjugated in the past. Finally, Yang and Shu (2016) performed a meta-analysis on a large number of fMRI experiments where subjects were listening to literal action sentences, fictive motion sentences, metaphorical action sentences, and idioms, and found increased activation in motor regions during metaphorical action sentences. This activation is thought to contribute to understanding and mapping metaphors onto their concrete reference. Simulation theories are often associated with embodied cognition, proposing that we use our brains and bodies to ground conceptual and abstract content in sensorimotor systems.

The third theory we discuss is sensory prediction. In this case what is being coded for in the motor system is sensorimotor prediction, or continuous online prediction of the very next state of the stimulus. A predictive role for the motor system is suggested in Clark's (2015) theory of "embodied prediction", in which motor activation during action observation would entail prediction of the upcoming sensory signal based on the current sensory information. In this case, motor cortex would be active during the observation of a grasping action, because the observer's brain would be actively predicting the very next movement via motor regions. Thus, a predictive role can account for the findings from the action-observation network literature. Wilson and Knoblich (2005) outline a version of the perceptual prediction role of motor areas that uses what they call perceptual emulators. An emulator is a mental simulation that models the external world. The emulator continues updating the model online, and the output from this emulator can be compared to the actual state of the external world to verify that expectations are being met. Emulators running in the motor system would internalize a model of the

biomechanics of the human body, allowing observers to model the movements of an observed agent as they unfold in time. Importantly, these emulators are running one step ahead of sensory input, predicting the upcoming external state before it happens and then comparing real to modeled state afterward.

If the motor system uses these emulators, it should also be able to model predictable sensory information that isn't human-made, such as rhythmic waves or the bouncing of a ball, by internalizing a model for the observed system. Supporting research comes from Schubotz (2007), whose work suggests that even observation of movements coming from non-animate entities recruits the motor system. In a number of fMRI experiments, they find that particular types of perceptual prediction tasks involving such things as pitch identification, spatial or object-related identification tasks activates premotor areas in a somatotopic way, similar to effector-related observation/execution tasks. For instance, object-related tasks recruited regions of premotor cortex that share activation in hand-related execution and observation tasks. As there isn't a common repertoire to humans and rolling waves, these findings could not be explained under the motor resonance account.

We propose that the role of the motor system varies depending on context. For instance, during the perception of action language, the motor system might serve to provide the motor component of covert simulation that occurs in the embodied processing of language. During the observation of very well practiced movements by an outside actor, the motor system may serve the purpose of driving motor resonance in the observer to quickly map the actions to the observer's body and understand the action. Finally, during perceptual processing of non-human movements, the motor system might serve to assist in perceptual processing by way of predictive emulator models.

One potential way of differentiating between prediction and simulation is by observing how the modulation of the motor system changes over repetitive viewing of stimuli. If the observer is simulating the observed action, then we should see a steady facilitation of MEPs across repetitions of a stimulus, indicating simulation at each occurrence. If motor system facilitation is due to predictive processes, however, might expect a different pattern of modulation. Because less error correction takes place as a stimulus becomes more predictable, we can predict that the neural populations underlying the predictive processes will be less active for more predictable sensory stimuli. Thus, we should see a decrease in corticospinal excitability over multiple repetitions of a stimulus, as it becomes more predictable and leads to lower error correction.

In this experiment we look at corticospinal excitability using single-pulse transcranial magnetic stimulation (TMS) during the perception of written language to examine the extent of motor involvement in a few variations of the stimuli. Subjects viewed videos of words being written out with a stylus and of words being typed letter by letter. Previous analyses in our lab have shown that observation of handwriting leads to motor simulation, while observation of typed words does not. We proposed that this is because while it is apparent that the handwritten text are human created, this is less apparent for text created on a keyboard. We repeat all stimuli four times over the course of the experiment. We predicted that MEPs in the handwritten stimuli trials would show an even facilitation across all presentations of the stimuli, because simulation should be consistent no matter how predictable it is. We hypothesized that the MEPs in the typed

stimuli trials would show initial facilitation, which would lessen as the stimuli are repeated and there is less prediction error. This would be expected because the first presentations of the stimuli, appearing letter by letter, should be difficult to predict, resulting in a large error in perceptual prediction. As stimuli are presented more often, perceptual prediction should become easier and less effort required on the part of error correction.

Methods

Participants

Twenty-four right-handed normal participants (8 males, 16 females, mean age ~ 19.5) were recruited in this study through UC Merced's SONA research system. All participants passed a safety screen and gave written, informed consent. The experimental procedure was approved by the UC Merced Institutional Review Board. Participants received 2 research credits that can be used for credit in some undergraduate courses.

TMS and EMG recording

Corticospinal excitability was measured by the amplitude of motor evoked potentials (MEPs) recorded using electromyography (EMG) on the first dorsal interosseus (FDI) muscle of the right hand. Two small adhesive electrodes (1cm^2) were placed over the belly of the recorded muscle and a ground electrode was placed over a bone on the participant's elbow. A bandpass filter (50 Hz1,000 Hz) was applied to the EMG signal, which was digitized at 1,000 Hz for offline analysis. MEPs were elicited by applying single-pulse TMS to the FDI region of the left motor cortex. Pulses were delivered using a Magstim Rapid² with an attached 70-mm figure-of-eight coil positioned over the optimal scalp location with the handle pointing backward at 45 degrees from the midline. The procedure was as follows. Subjects were fitted with a swim cap that was covered by a grid of dots placed 1 cm² apart. Optimal scalp position was determined by moving the coil by one centimeter intervals until the location eliciting the best MEPs was identified. This location was marked on the swim cap worn by the participant. Resting motor threshold was determined as the percent of machine output that produced 5 out of 10 MEPs of at least 50 μV peak-topeak amplitude. The stimulation intensity during the experiment was set to 120% of a participant's resting motor threshold. The coil was held steady at the optimal position throughout the experiment. Subjects were instructed to keep their head still and remain relaxed for the duration of the experiment.

Experimental paradigm

The visual stimuli consisted of videos of either handwritten or typed words or nonwords appearing letter by letter at a variable typing speed averaging 3-4 letters per second. Words were chosen that did not relate to any actions or manipulable objects, to ensure that our measurement would not be influenced by the effects of semantic processing of action. We also included 10 baseline trials, which consisted of a single black box for the same duration as the stimuli. We chose to randomize the baseline trials in with the rest of the trials so that the baseline measure would not be biased by a lack of attention that can occur when baseline measures are all recorded pre-experiment. Stimuli included ten linguistic stimuli, which appeared four times in each of the conditions. This

resulted in 80 stimuli trials and 10 baseline trials, or a total of 90 trials. Eight seconds passed in between individual trials, and the total experiment length was approximately 12 minutes. We chose to apply stimulation two seconds into the ongoing video, so that as the stimuli were repeated, they were more highly predictable (by the presence of the first few letters) by the time stimulation occurred. Because TMS stimulation would occur two seconds into the video, we ensured that the typed stimuli would display one of the following letters at that time [N, H, U, M, J, I], so that if subjects were simulating the typing in proper typing position, FDI would be the simulated muscle.

The stimuli appeared on a computer screen in front of the participants. Participants were instructed to attend to the stimuli on the screen and were given notice when the experiment was one-third and two-thirds of the way finished to prevent loss of attention. TMS pulses were delivered 2 seconds after video onset. The interval between trials was 8 seconds, to avoid any cumulative effects of single-pulse TMS. After the experiment, subjects were asked whether they were able to stay attentive during the length of the experiment. Participants who said they were not were excluded from analyses (5 subjects).

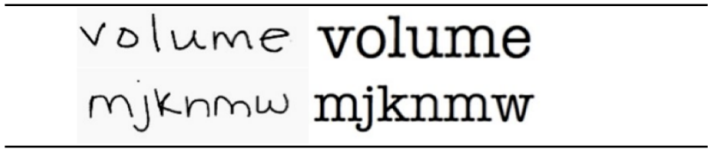


Figure 1: Examples of stimuli used in the experiment. Handwritten stimuli are on the left and typed stimuli on the right. In the experiment, participants saw the stimuli appear as a video as they were written or typed.

Results

The average raw MEP amplitude for handwritten stimuli was 1.126 mV, with a standard deviation of 1,303. The average for typed words was 1098 mV, with a standard deviation of 1.295. Because of the large variations between participants, raw MEP amplitude values were z-scored to allow inter-individual comparisons. The resulting z-scores indicate the distance (in standard deviations) that a particular MEP score is from the mean. Figure 2 shows the average zscore in each condition. The average z-score for handwritten stimuli was .1, while that for typed stimuli was -.06.

A two-way repeated-measures analysis of variance (ANOVA) was computed on the standardized MEPs to test for significant effects. The considered factors were condition (handwritten or typed) by order (nth time that a stimulus appeared). We observed a significant main effect for condition, with handwritten stimuli producing greater facilitation of MEP amplitude with respect to typed stimuli, $F(1,23) = 7.62$, $p < .01$. We also observed a significant interaction effect of condition by order of presentation, $F(3,184) = 3.77$, $p = .05$. In particular, there was a consistent facilitation in MEPs in the handwritten stimuli regardless of how many times the stimulus has been presented. In the typed stimulus condition, however, there was an initial facilitation in the MEP amplitude that decreased with each repetition of the stimuli. This pattern of results confirms our hypothesis of typed stimuli showing an initial facilitation of corticospinal

excitability, followed by a decrease in that facilitation. This also confirms our hypothesis that the handwritten words would induce consistent facilitation of corticospinal excitability.

A linear regression of presentation number on baseline zscore was performed in order to evaluate whether the baseline MEPs changed with multiple presentations of the stimuli. The regression came out non-significant ($t = -1.1$, $p > .3$). This indicates that overall MEP amplitudes are not varying as a function of time or number of repetitions to stimuli.

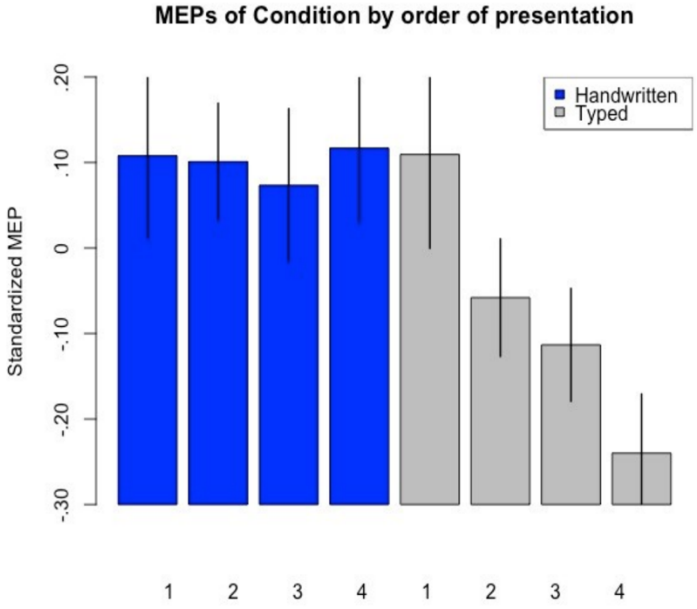


Figure 2: Standardized (Z-scored) MEP amplitudes for each condition. X-axis shows presentation number (nth time a stimulus was presented). Motor evoked potentials in the handwritten condition show consistent facilitation, while those in the typed condition show initial facilitation that decreases with presentation number.

Discussion

In this experiment we observed a differential pattern of motor facilitation dependent on word reading condition. In particular, the observation of actively handwritten words produced a persistent facilitation in MEP amplitudes. This is consistent with the action observation research, where subjects view actions produced by others over multiple trials and produce consistent MEP facilitation. When subjects are exposed to actively typed words, however, the pattern of MEP facilitation changes, with repetitive exposure to the stimulus resulting in a decrease in observed corticospinal excitability. In previous work, we hypothesized that typed stimuli might not show simulation because of two reasons. Either the act of typing has weak or no sensorimotor association, or the discrete nature of typed words does not invoke simulation the same as the continuous strokes of handwriting.

Evidently the motor system is doing something different from motor simulation during the observation of words that are actively typed. One potential hypothesis is that corticospinal excitability in the typed condition is influenced by attentiveness. As subjects are repeatedly exposed to words, they might lose interest and thus exhibit lower attention. We included in the experiment a baseline measure appearing randomly throughout, consisting of a solid black box that appears instead of the language stimuli. There were 10 baseline trials used. If corticospinal excitability was picking up on a measure of attention, we should see a predictable decreasing trend in MEP amplitudes across repetitions of the baseline trials as well. No such decreasing trend was observed over the repeated baseline trials. Though we cannot rule out the possibility entirely, this does suggest that there is something happening for the typed stimuli other than decreased attentiveness.

We suggest that the decrease in excitability across repetitions of stimuli is due to sensory prediction by the motor system. When the stimuli are less predictable (i.e., the first presentations), the sensory prediction error is large, resulting in higher motor activation. As the stimuli are repeated and become more predictable, the sensory prediction error becomes lower and we observe less corticospinal excitability in the motor system. This account is consistent with Schubotz's (2007) findings of motor activation during serial prediction tasks and Wilson and Knoblich's (2005) emulator account.

If our theoretical formulation is correct, this implies that the study of motor involvement in perception and cognition should take into account that the motor system is playing multiple processing roles that are network and context dependent. The action observation based recruitment of the motor system is well established. Strong evidence suggests that this is due to motor resonance that is both effector specific and time-dependent. We contend that the role of motor cortex in action-observation is for low-level activation of one's own motor repertoire. Under our account, motor activation during perceptual processing of nonhuman-created stimuli, reported by Schubotz and colleagues, is not at odds with the resonance account of action observation. The particular information processing role of motor regions does not need to be identical across contexts. The functional network underlying action observation includes bilateral mid-temporal gyrus (MTG) and left inferior parietal lobule as well as left premotor cortex. (Gazzola, Aziz-Zadeh, & Keysers, 2006). Other brain regions active during figurative language include the left and right inferior frontal gyrus (IFG), bilateral medial frontal gyri (medFG), left temporal lobe, and amygdala. (Bohrn, Altmann, & Jacobs, 2012). The function of motor activation in each of these different networks can be defined by its connections and interactions, allowing a motor predictive system or motor simulation system when appropriate.

How would this region have multiple functional roles? Evidence from single-unit recording of neurons in premotor areas suggests that there is a wide variety of neurons that respond to different contexts. For example, during the discovery of mirror neurons, many types of such neurons were identified (Di Pellegrino et al., 1992). Some of these are called "strictly congruent" mirror neurons, which respond to action observation and action execution only to the same exact movement. More common were "broadly congruent" mirror neurons, which respond to action observation and action execution

during similar types of movements, encompassing a broader response range. We postulate that the first type is responsible for driving motor resonance-related activation, while the latter type could potentially underlie the sort of sensory prediction we discuss. Finally, a third type of neuron they observed was called a “canonical neuron”, which respond to the observation of manipulable objects. Perhaps these neurons could play a role in mental simulation, or affordance processing. These examples are all speculative and not grounded by any evidence in the present work, but they aim to push intuitions toward a fresh perspective. Future work using single-neuron recording would be needed to directly test such hypotheses. At a brain region level, however, we can learn more by observing how activation in local regions changes with repetition of sensory stimuli or changes in stimuli.

Future research that we are currently engaged aims to explore how sensorimotor contingencies are learned by training participants on novel sensory to motor mappings. We will then use these controlled artificial mappings to explore sequential prediction and/or simulation using the motor system.

Chapter 7
General Discussion

1. Motor areas of the brain are important for many perceptual and cognitive processes

The motor system was once thought of only as a network of brain regions responsible for controlling the movements of our bodies. Substantial evidence of motor system involvement during motor imagery and action observation expanded our vision of the role of this network, by suggesting that action simulation is a form of action itself (Jeanerrod, 2001). We now have evidence that these brain regions are important for yet more cognitive and perceptual processes.

I have laid out examples of three different experiments providing evidence for motor system involvement in perception. First, I have shown that motor system activity increases during the passive observation of handwritten words where no visible agent is creating the simulated movements (see Chapter 2). I have also shown that a group of motor regions is consistently active while listening to music even in non-musicians who do not have the experience to simulate the performance of the musical pieces (see Chapter 3). Finally, I found evidence that motor system activation during observation of piano playing can be more or less predictive depending on the actions observed (see Chapter 5).

Other cognitive processes have been shown to involve the motor system as well. Timing and entrainment to external stimuli engage the motor system (Ross & Balasubramaniam, 2014) particularly in musical beat perception (Ross, Iversen, & Balasubramaniam, 2016). Prediction of events unfolding in time recruit functionally related somatotopic regions of lateral vPMC (Schubotz, 2007). Specifically, hand-related areas of vPMC are active while predicting object size while mouth-related regions are active during prediction of pitch modulation. These findings in combination with the studies presented in this dissertation strongly suggest an additional role for this brain network.

2. Predictive motor system perception

Predictive processing theories propose that the primary function of the brain is to reduce uncertainty by actively predicting upcoming events (Clark, 2015; Friston, Kilner, & Harrison, 2006). We learn predictive models of the systems that we interact with, and these models are improved upon with more experience. In this vein, the motor system is a collection of neural networks whose circuitry is adapted for predictive processing of motor dynamics (Jeannerod, 2001; Wolport & Flanagan, 2001). The computations needed for predictive motor control when we move must integrate information about continuous dynamics of the body in space and time across multiple sensory modalities. The motor brain regions are exapted for predicting and modeling other perceptual and cognitive processes using these same computational resources. This process has also been referred to as a form of emulation (Grush, 2004). The more experience that individuals have with any given system or action, the better their predictive models of the events become, leading to faster processing and better inferences regarding the state of the system.

Different regions within the motor network may be specialized for different kinds of perceptual prediction. Attending to different aspects of a predicted auditory stimulus differentially engages parts of premotor cortex (Schubotz, von Cramon, & Lohmann, 2003). Specifically, attending to "where" auditory events took place engages parts of dorsolateral premotor cortex important for gaze and reaching. Attending to "what" auditory events took place engages superior ventrolateral premotor cortex, crucially involved in hand movements. Lastly, attending to "when" auditory events occur activates inferior ventrolateral premotor cortex, responsible for controlling mouth movements. Thus, different parts of premotor cortex are best suited to different prediction-related computations. This pattern of somatotopic neural reuse may hold generally across the motor system as the predictive substrate of additional perceptual and cognitive domains.

3. Simulation and the motor system

Mental simulation is the process of internally generating sensorimotor information that is distinct from or missing in the environment. Some theories suggest that, during action observation, motor simulation occurs automatically through motor resonance. In this view, we automatically simulate others' actions with our own bodies. This process is akin to that during motor imagery and Jeannerod suggests that the only difference between these mechanisms and overt actions is that in the former, movement is inhibited either because the neural signal is not strong enough or an inhibition prevents the movement from occurring. In cases where a goal can be inferred from an observed action, motor simulation explains the recruitment of motor regions during perceptual processing. We use simulations of these observed actions to understand the goals and intentions of actors by reactivating goals and intentions that we have experienced while performing that action. Simulations can also facilitate the recruitment of other sensory modalities, such as the auditory consequences of a simulated movement.

Many of the motor theories of perception (see Chapter 1) use the term "simulation" to describe what happens during perceptual recruitment of motor areas. As a result, simulation is used to describe a broad variety of phenomena across a variety of disciplines, often without specification of what is meant in terms of the cognitive, experiential, and neural implementation of simulative processes. I propose that mental simulation is best viewed as a cognitive tool that builds on top of predictive processing mechanisms that are continuously operating across modalities. Mental simulation adds to general prediction mechanisms. It enriches these predictions with the experience of the simulated events and actions. It is important to lay out what that simulation entails and outline the ways that this notion of simulation differs from others in various theoretical accounts.

The predictive role of motor areas in perceptual processing can explain the ubiquitous activation of those regions during what appear to be basic perceptual tasks. In this view, mental simulation is a tool that is only used in situations where the environment actively cues sensorimotor experiences and invoking these representations is useful. It is not the case that whenever we see motor activation during perception, we must posit a simulation process, as the role of these regions in perceptual prediction can give rise to this activation. Beat perception is one example of such a perceptual process that likely exapts the predictive role of motor regions but does not invoke motor simulation. In this case,

there is generally no direct entity or action to simulate, only a pattern to predict. The experiments described above regarding pitch perception and object size prediction are also examples of processes that take advantage of the predictive capacity of the motor circuitry without clear evidence of mental simulation. In my view, these processes emulate the predicted systems by extending the features of pitch production models and grasping models to these events. Emulators extend features of a similar, analogous system to the observed system, allowing ongoing prediction to occur and predictive models of the system to build up. For example, the neural substrate responsible for predicting the movement of one's arms might be computationally suitable for the prediction of waves rolling on the ocean and thus be recruited to emulate this system (Schubotz, 2007).

The cognitive tool view might define mental simulation in terms of three important axes that can separate different kinds of simulation. The first axis is explicit/implicit, illustrating the difference between mentally simulating rotating a shape intentionally to solve a puzzle, and implicit mirroring. The second is bottom-up (reflexive)/top-down (deliberate), which differentiates sensory-driven and top-down predictive simulation. Finally, the third axis is 1st person/3rd person, defined by whether the person doing the simulating is enacting this process from one's own perspective, or an outward perspective. The different locations in this 3-dimensional space may correspond to differences in cognitive and neural implementation of the simulation process. For instance, motor simulation that involves the reenactment of one's own bodily state during an observed action to predict and understand that action resides in the first person, implicit, bottom-up position on these axes. See Figure 7.1 for a depiction of these axes populated with some domains that reference simulation.

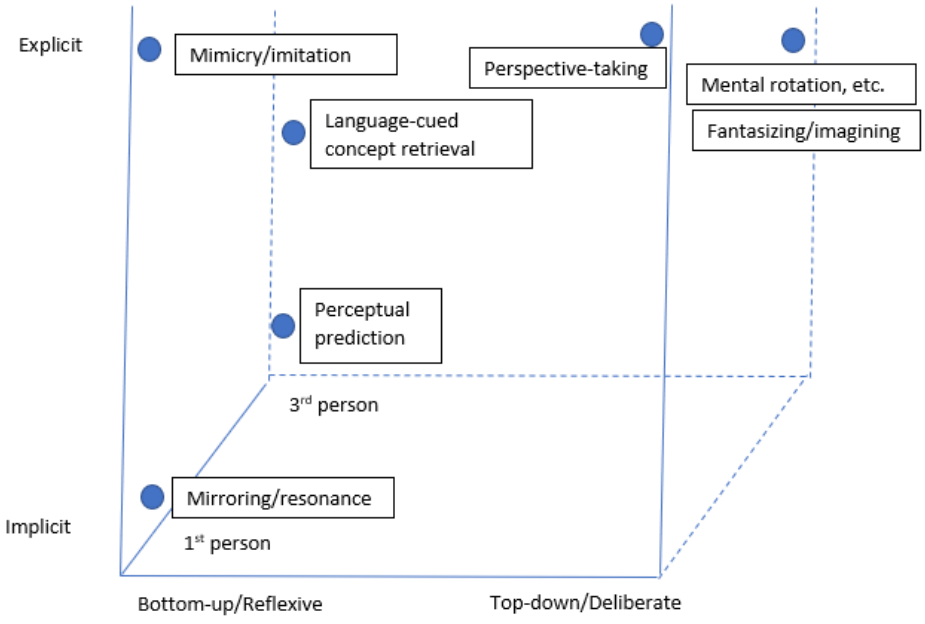


Figure 7.1 Depiction of the proposed simulation axes

Mental simulation engages the respective areas of the brain that would be active if performing a simulated task. Imagining oneself performing an action consists of activation of the sensorimotor circuits responsible for performing the action. These circuits can include information about actual movement execution, such as movement trajectory, but they can also include other experiential factors that are typically present when performing that action. For instance, if I am picking up a cup, this action corresponds to my feeling thirst. When I observe another individual picking up a cup, mental simulation can activate the respective brain areas that are active when I pick up a cup, including those that process information about thirst. This is how we can infer the goals, intentions, and feelings of others through mental simulation.

4. Full and partial simulation

Given the position that motor simulation is a cognitive tool that is sometimes applied during perception, ongoing motor simulations need not supplement all perceptual processes. Mental simulation also varies in terms of isomorphism with the simulated action or experience. For instance, simulating the trajectory of a projectile so that one can accurately predict where it will land might invoke an isomorphic simulative process (third person, explicit, bottom-up). In other words, the simulation will unfold mentally in the same temporal units that the simulated event unfolds in. We can call this a "full simulation". When imagining an event such as a bike ride you will take later, however, you might not simulate all of the features of the action (taking your bike out of the house, swinging your leg over the top of the bike, and so forth). It is more likely that your imagination of this future event will involve a simulation that takes shortcuts and abstracts away from less important implementation details of the action or event. If you imagine riding down a familiar street, this imagery will not take place over the same length of time as it takes to actually ride your bike down that street.

This ability to perform abstracted simulations of past or future events is what allows us to be creatures that can remember the past and plan for our futures. This may also be important for how we represent and learn concepts. Barsalou (2009) describes concept activation as happening through the partial reactivation of the patterns of brain activation previously experiences with the concept. We can think of these as mental simulations that are yet further abstracted from full temporal simulation of an event. For example, if the concept "car" comes up during a conversation, this might reactivate some relevant sensory representations associated with cars, such as the visual representation of my car and the posture and movements associated with driving my car (first person, implicit, top-down). I will likely not, however, recruit a temporally extended mental simulation of driving my car. In other words, concepts and knowledge can be represented through very high-order, non-isomorphic simulation, which is just to say they can be represented through partial reactivation of the relevant brain states, which we can call "partial simulation".

Affordances, or the direct interactions we can have with our external environment, are an important feature of object representations. I have shown that affordances have direct consequences for how objects are mentally represented (see Chapter 4). I would further suggest that affordances are one of the key features of mental simulation of objects. This means that when someone mentions an object like a "cup", the partial reactivation

constituting the simulation includes the affordances for cup such as “can be picked up” and “can be drunk from”. Top-down factors such as context can further constrain which aspects of the event or object receive partial re-activation.

5. Learning motor skills changes perceptual processing

Learning a new skill such as playing an instrument will make it easier to apply mental simulation in the related sensorimotor domain. For example, if a non-musician hears a musical piece, they will automatically apply sensorimotor prediction as the sound unfolds. If a drummer listens to the same piece, they might actively simulate the drumming actions that produce the beat of the song. This additionally changes the way that musicians perceive music by changing the top-down processes of attending to features of the music itself. It remains to be explored how skill learning changes predictive mechanisms during perception and performance of the involved actions.

Observational learning has been shown to result in some motor learning as well, suggesting that it is not necessary to have direct experience with a skill in order to acquire knowledge of the involved motions. By watching others perform skilled movements, we can mirror those movements with our own bodies and build up an impoverished model of that skill. This experience of observing others do something we have not done also allows us to simulate those actions offline or while receiving partial sensory information about these events in real time.

6. Motor simulation theories: overlap and divergence

I have discussed numerous theories for motor involvement in the different domains of perception, from speech to music to human action perception. Some of the distinctions between these theories that are not readily apparent may be made clearer through the axes model of simulation. The motor theory of speech perception, for instance, would be described as first-person, implicit, bottom-up simulation. ASAP is summarized as a third person, implicit, bottom-up simulation. Perhaps a key defining difference between theories of the former variety and of the latter variety is whether there is a direct mapping of the perceived stimulus to one’s own body. Mirror neuron theories typically fall into the same category as the motor theory of speech perception, but when mirroring gives way to actual imitation, this becomes explicit rather than implicit. Grush’s emulation theory (as well as the HAPEM framework) can be described as third-person, implicit, and top-down simulation, and helps to describe how the motor system can be recruited for non-action events.

The goal of determining a unified theory of recruitment of motor areas for perception, I believe, is not the right goal. As we have seen, context and experience changes the neural processes underlying many perceptual experiences. Thus, theories of motor prediction and theories of mental simulation can both offer sound explanations for motor system involvement in different perceptual tasks.

In the experiments I described, I point to motor system activation tied to motor simulation or prediction. When handwritten words are observed on a screen, this taps into a simulation mechanism as the ongoing dynamics of the handwriting lends itself to simulating the writing act itself. When typed words are observed, motor system activity was increased as well, but more so when seeing a word for the first or second time. I

proposed that this reflects a predictive process occurring, partially in motor areas, which will induce less necessary neural activation of the prediction system as prediction errors are reduced. As each typed letter appears instantaneously, and is not temporally drawn out, simulation is not afforded by these words. I also showed that listening to music activates primary and secondary motor areas, as well as cerebellum. These are important areas for motor prediction and are likely recruited for sensory prediction. Lastly, I showed that watching someone play the piano engages motor regions. This is an event that will recruit simulation, as the observed ongoing actions can be simulated with the observer's body. Interestingly, motor activity appears to display a predictive pattern of encoding during movements that are easy and common, but not during movements that are harder to make and less frequently executed. This may mean that actions without good predictive models are more likely to recruit a temporally extended simulation of the action.

In summary, a vast literature points to the motor system proper not being a neural network that is only good for controlling and planning our actions. As we develop the vocabulary of the field to use terms like "action-perception loops" and discuss these processes as less separable than previously considered, perhaps we should also reconsider the term "motor system" to reflect its diverse roles in sensorimotor prediction.

References

Alluri V, Toiviainen P, Jääskeläinen IP, Glerean E, Sams M, Brattico E. Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm. Neuroimage. 2012; doi:10.1016/j.neuroimage.2011.11.019

Anderson ML. Précis of after phrenology: Neural reuse and the interactive brain. Behav Brain Sci. 2015;39. doi:10.1017/S0140525X15000631

Aziz-Zadeh, L., Iacoboni, M., Zaidel, E., Wilson, S., & Mazziotta, J. (2004). Left hemisphere motor facilitation in response to manual action sounds. *European Journal of Neuroscience*, *19*(9), 2609-2612.

Aziz-Zadeh, L., Koski, L., Zaidel, E., Mazziotta, J. & Iacoboni, M. (2006a) Lateralization of the human mirror neuron system. Journal of Neuroscience 26 (11):2964–7

Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. Philosophical Transactions of the Royal Society of London B: Biological Sciences, 364(1521), 1281-1289.

Baumann S, Koeneke S, Schmidt CF, Meyer M, Lutz K, Jancke L. A network for audio-motor coordination in skilled pianists and non-musicians. Brain Res. 2007;1161: 65–78. doi:10.1016/j.brainres.2007.05.045

Bangert M., Altenmüller E. O. (2003). Mapping perception to action in piano practice: a longitudinal DC-EEG study. *BMC Neurosci.* 4:26. 10.1186/1471-2202-4-26

Bangert M, Peschel T, Schlaug G, Rotte M, Drescher D, Hinrichs H, et al. Shared networks for auditory and motor processing in professional pianists: Evidence from fMRI conjunction. Neuroimage. 2006;30: 917–926. doi:10.1016/j.neuroimage.2005.10.044

Barsalou, L. W. (1999). Perceptions of perceptual symbols. *Behavioral and brain sciences*, *22*(4), 637-660.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617-645.

Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1281-1289.

Baumgartner T, Lutz K, Schmidt CF, Jäncke L. The emotional power of music: How music enhances the feeling of affective pictures. Brain Res. 2006; doi:10.1016/j.brainres.2005.12.065

Beilock, S. L., & Holt, L. E. (2007). Embodied preference judgments: Can likeability be driven by the motor system?. *Psychological Science*, *18*(1), 51-57.

Bengtsson SL, Ullén F, Henrik Ehrsson H, Hashimoto T, Kito T, Naito E, et al. Listening to rhythms activates motor and premotor cortices. Cortex. 2009;45: 62–71. doi:10.1016/j.cortex.2008.07.002

Bestmann S., Harrison L. M., Blankenburg F., Mars R. B., Haggard P., Friston K. J., et al. (2008). Influence of uncertainty and surprise on human corticospinal excitability during preparation for action. *Curr. Biol.* 18 775–780. 10.1016/j.cub.2008.04.051

Blood, A., Zatorre, R., Evans, A. Intensely pleasant emotional responses to music correlate with CBF modulation in paralimbic and other subcortical brain regions. Physiol Psychol. 1996;8: 126–9.

Bohrn, I. C., Altmann, U., & Jacobs, A. M. (2012). Looking at the brains behind figurative language—A quantitative meta-analysis of neuroimaging studies on metaphor, idiom, and irony processing. Neuropsychologia, 50(11), 2669-2683.

Brincker, M. (2010). *Moving beyond mirroring-A social affordance model of sensorimotor integration during action perception*. City University of New York.

Brown S, Martinez MJ. Activation of premotor vocal areas during musical discrimination. Brain Cogn. 2007;63: 59–69. doi:10.1016/j.bandc.2006.08.006

Brown S, Martinez MJ, Parsons LM. Passive music listening spontaneously engages limbic and paralimbic systems. Neuroreport. 2004;15: 2033–2037. doi:10.1097/00001756-200409150-00008

Bruderer, A. G., Danielson, D. K., Kandhadai, P., & Werker, J. F. (2015). Sensorimotor influences on speech perception in infancy. *Proceedings of the National Academy of Sciences*, *112*(44), 13531-13536.

Buccino, G., Vogt, S., Ritzl, A., Fink, G. R., Zilles, K., Freund, H. J. & Rizzolatti, G. (2004) Neural circuits underlying imitation learning of hand actions: An eventrelated fMRI study. Neuron 42(2):323–34

Buccino, G., Riggio, L., Melli, G., Binkofski, F., Gallese, V., & Rizzolatti, G. (2005). Listening to action-related sentences modulates the activity of the motor system: A combined TMS and behavioral study. *Cognitive Brain Research*, *24*(3), 355-363.

Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V., ... & Freund, H. J. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *European journal of neuroscience*, *13*(2), 400-404.

Calvo-Merino, B., Glaser, D. E., Grèzes, J., Passingham, R. E., & Haggard, P. (2005). Action observation and acquired motor skills: an FMRI study with expert dancers. *Cerebral cortex*, *15*(8), 1243-1249.

Calvo-Merino, B., Grèzes, J., Glaser, D. E., Passingham, R. E., & Haggard, P. (2006). Seeing or doing? Influence of visual and motor familiarity in action observation. *Current Biology*, *16*(19), 1905-1910.

Candidi, M., Leone-Fernandez, B., Barber, H. A., Carreiras, M., & Aglioti, S. M. (2010). Hands on the future: facilitation of cortico-spinal hand-representation when reading the future tense of hand-related action verbs. European Journal of Neuroscience, 32(4), 677-683

Candidi M., Maria Sacheli L., Mega I., Aglioti S. M. (2012). Somatotopic mapping of piano fingering errors in sensorimotor experts: TMS studies in pianists and visually trained musically naives. *Cereb. Cortex* 24 435–443. 10.1093/cercor/bhs325

Caria A, Venuti P, De Falco S. Functional and dysfunctional brain circuits underlying emotional processing of music in autism spectrum disorders. Cereb Cortex. 2011; doi:10.1093/cercor/bhr084

Casasanto, D., & Boroditsky, L. (2008). Time in the mind: Using space to think about time. *Cognition, 106*(2), 579-593.

Chen JL, Penhune VB, Zatorre RJ. Listening to musical rhythms recruits motor regions of the brain. Cereb Cortex. 2008;18: 2844–2854. doi:10.1093/cercor/bhn042

Clark, A. (2015). Embodied prediction. In *Open Mind*. Open MIND. Frankfurt am Main: MIND Group.

Cook, R., Bird, G., Catmur, C., Press, C., & Heyes, C. (2014). Mirror neurons: from origin to function. *Behavioral and Brain Sciences*, *37*(02), 177-192.

Csibra, G. (2005). Mirror neurons and action observation: Is simulation involved. *What do mirror neurons mean*.

Cong F, Puoliväli T, Alluri V, Sipola T, Burunat I, Toiviainen P, et al. Key issues in decomposing fMRI during naturalistic and continuous music experience with independent component analysis. J Neurosci Methods. 2014; doi:10.1016/j.jneumeth.2013.11.025

D'Ausilio A., Altenmüller E., Olivetti Belardinelli M., Lotze M. (2006). Cross-modal plasticity of the motor cortex while listening to a rehearsed musical piece. *Eur. J. Neurosci.* 24 955–958. 10.1111/j.1460-9568.2006.04960.x

Demorest SM, Morrison SJ, Stambaugh LA, Beken M, Richards TL, Johnson C. An fMRI investigation of the cultural specificity of music memory. Soc Cogn Affect Neurosci. 2009;5: 282–291. doi:10.1093/scan/nsp048

De Nooijer, J. A., Van Gog, T., Paas, F., & Zwaan, R. A. (2013). When left is not right: handedness effects on learning object-manipulation words using pictures with left-or right-handed first-person perspectives. *Psychological Science, 24*(12), 2515-2521.

di Pellegrino, G. D., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental brain research*, *91*(1), 176-180.

Dobek CE, Beynon ME, Bosma RL, Stroman PW. Music modulation of pain perception and pain-related activity in the brain, brain stem, and spinal cord: A functional magnetic resonance imaging study. J Pain. 2014; doi:10.1016/j.jpain.2014.07.006

Doupe, A. J., & Kuhl, P. K. (1999). Birdsong and human speech: common themes and mechanisms. *Annual review of neuroscience*, *22*(1), 567-631.

Dushanova J., Donoghue J. (2010). Neurons in primary motor cortex engaged during action observation. *Eur. J. Neurosci.* 31 386–398. 10.1111/j.1460-9568.2009.07067.x

Eickhoff S, Laird A, Grefkes C, Wang LE, Zilles K, Fox PT. Coordinate-based ALE meta-analysis of neuroimaging data: a random-effects approach based on empirical estimates of spatial uncertainty. Hum Brain Mapp. 2009;30: 2907–2926. doi:10.1002/hbm.20718

Ertelt, D., Small, S., Solodkin, A., Dettmers, C., McNamara, A., Binkofski, F., & Buccino, G. (2007). Action observation has a positive impact on rehabilitation of motor deficits after stroke. *Neuroimage*, *36*, T164-T173.

Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *European Journal of Neuroscience*, *15*(2), 399-402.

Fadiga L, Craighero L, Olivier E. Human motor cortex excitability during the perception of others' action. Curr Opin Neurobiol. 2005;15: 213–218. doi:10.1016/j.conb.2005.03.013

Fadiga L, Fogassi L. Motor facilitation during action observation: a magnetic stimulation study. J Neurophysiol. 1995;73: 2608–2611. doi:10.1.1.299.4524

Fang J, Hu X, Han J, Jiang X, Zhu D, Guo L, et al. Data-driven analysis of functional brain interactions during free listening to music and speech. Brain Imaging Behav. 2015; doi:10.1007/s11682-014-9293-0

Ferrari P. F., Gallese V., Rizzolatti G., Fogassi L. (2003). Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex.*Eur. J. Neurosci.* 17 1703–1714. 10.1046/j.1460-9568.2003.02601.x

Fischer, M. H., & Zwaan, R. A. (2008). Embodied language: A review of the role of the motor system in language comprehension. *The Quarterly Journal of Experimental Psychology*, *61*(6), 825-850.

Flores-Gutiérrez EO, Díaz JL, Barrios FA, Favila-Humara R, Guevara MÁ, del Río-Portilla Y, et al. Metabolic and electric brain patterns during pleasant and unpleasant emotions induced by music masterpieces. Int J Psychophysiol. 2007; doi:10.1016/j.ijpsycho.2007.03.004

Fodor JA. Fodor's guide to mental representation: The intelligent auntie's vade-mecum. Mind. 1985;94: 76–100

Fogassi L., Ferrari P. F., Gesierich B., Rozzi S., Chersi F., Rizzolatti G. (2005). Parietal lobe: from action organization to intention understanding. *Science* 308 662–667. 10.1126/science.1106138

Friston, K., Kilner, J., & Harrison, L. (2006). A free energy principle for the brain. *Journal of Physiology-Paris*, *100*(1-3), 70-87.

Friston K, Mattout J, Kilner J. Action understanding and active inference. Biol Cybern. 2011;104: 137–160. doi:10.1007/s00422-011-0424-z

Gallese, V. (2007). Before and below 'theory of mind': embodied simulation and the neural correlates of social cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1480), 659-669.

Gallese V, Lakoff G. The brain's concepts: The role of the sensory-motor system in conceptual knowledge. Cogn Neuropsychol. 2005;22: 455–479. doi:10.1080/02643290442000310

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, *119*(2), 593-609.

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in cognitive sciences*, *2*(12), 493-501.

Gallese V. (2008). Mirror neurons and the social nature of language: the neural exploitation hypothesis. *Soc. Neurosci.* 3 317–333. 10.1080/17470910701563608

Gangitano M., Mottaghy F. M., Pascual-Leone A. (2001). Phase-specific modulation of cortical motor output during movement observation. *Neuroreport* 12 1489–1492. 10.1097/00001756-200105250-00038

Gangitano M., Mottaghy F. M., Pascual-Leone A. (2004). Modulation of premotor mirror neuron activity during observation of unpredictable grasping movements. *Eur. J. Neurosci.* 20 2193–2202. 10.1111/j.1460-9568.2004.03655.x

Gaser C, Schlaug G. Brain structures differ between musicians and non-musicians. J Neurosci. 2003;23: 9240–9245. doi:23/27/9240

Gazzola, V., Aziz-Zadeh, L., & Keysers, C. (2006). Empathy and the somatotopic auditory mirror system in humans. Current Biology, 16(18), 1824-1829.

Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences*, *20*(1), 1-19.

Glenberg, A. M., Havas, D., Becker, R., & Rinck, M. (2005). Grounding language in bodily states: The case for emotion. In D. Pecher & R. A. Zwaan (Eds.), *Grounding*

Cognition: The Role of Perception and Action in Memory, Language, and Thinking. Cambridge University Press.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, *9*(3), 558-565.

González, J., Barros-Loscertales, A., Pulvermüller, F., Meseguer, V., Sanjuán, A., Belloch, V., & Ávila, C. (2006). Reading cinnamon activates olfactory brain regions. *Neuroimage, 32*(2), 906-912.

Grafton ST, Woods RP, Mazziotta JC, Phelps ME. Somatotopic mapping of the primary motor cortex in humans: activation studies with cerebral blood flow and positron emission tomography. JNeurophysiol. 1991;66: 735–743. doi:10.1152/jn.1991.66.3.735

Grahn JA. The Role of the Basal Ganglia in Beat Perception. Annals of the New York Academy of Sciences. 2009;1169: 35–45. doi:10.1111/j.1749-6632.2009.04553.x

Grahn JA, Brett M. Rhythm and beat perception in motor areas of the brain. J Cogn Neurosci. 2007;19: 893–906. doi:10.1162/jocn.2007.19.5.893

Grezes, J., & Decety, J. (2001). Functional anatomy of execution, mental simulation, observation, and verb generation of actions: a meta-analysis. *Human Brain Mapping*, *12*(1), 1-19.

Grèzes, J.,Armony,J. L.,Rowe,J.&Passingham,R.E.(2003) Activations related to "mirror" and "canonical" neurones in the human brain: An fMRI study. Neuroimage 18(4):928–37

Grush R. The emulation theory of representation: Motor control, imagery, and perception. Behav Brain Sci. 2004;27: 377–396. doi:10.1017/S0140525X04000093

Habermeyer B, Herdener M, Esposito F, Hilti CC, Klarhöfer M, Di Salle F, et al. Neural correlates of pre-attentive processing of pattern deviance in professional musicians. Hum Brain Mapp. 2009;30: 3736–3747. doi:10.1002/hbm.20802

Hald, L. A., van den Hurk, M., & Bekkering, H. (2015). Learning verbs more effectively through meaning congruent action animations. *Learning and Instruction, 39*, 107-122.

Halsband U, Ito N, Tanji J, Freund HJ. The role of premotor cortex and the supplementary motor area in the temporal control of movement in man. Brain. 1993;116: 243–266. doi:10.1093/brain/116.1.243

Hari R, Forss N, Avikainen S, Kirveskari E, Salenius S, Rizzolatti G. Activation of human primary motor cortex during action observation: a neuromagnetic study. Proc Natl Acad Sci U S A. 1998;95: 15061–5. doi:10.1073/pnas.95.25.15061

Harrington DL, Haaland KY. Neural Underpinnings of Temporal Processing: A Review of Focal Lesion, Pharmacological, and Functional Imaging Research. Reviews in the Neurosciences. 1999;10: 91–116. doi:10.1515/revneuro.1999.10.2.91

Harris R, De Jong BM. Differential parietal and temporal contributions to music perception in improvising and score-dependent musicians, an fMRI study. Brain Res. 2015; doi:10.1016/j.brainres.2015.06.050

Haslinger B., Erhard P., Altenmüller E., Schroeder U., Boecker H., Ceballos-Baumann A. O. (2005). Transmodal sensorimotor networks during action observation in professional pianists. *J. Cogn. Neurosci.* 17 282–293. 10.1162/0898929053124893

Haueisen J, Knösche TR. Involuntary motor activity in pianists evoked by music perception. J Cogn Neurosci. 2001;13: 786–792. doi:10.1162/08989290152541449

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, *41*(2), 301-307.

Hauk, O., & Pulvermüller, F. (2004). Neurophysiological distinction of action words in the fronto-central cortex. *Human Brain Mapping*, *21*(3), 191-201.

Havas, D. A., Glenberg, A. M., & Rinck, M. (2007). Emotion simulation during language comprehension. *Psychonomic Bulletin & Review, 14*(3), 436-441.

Heine L, Castro M, Martial C, Tillmann B, Laureys S, Perrin F. Exploration of functional connectivity during preferred music stimulation in patients with disorders of consciousness. Front Psychol. 2015; doi:10.3389/fpsyg.2015.01704

Hickok G, Poeppel D. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. Cognition. 2004;92: 67–99. doi:10.1016/j.cognition.2003.10.011

Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in cognitive sciences*, *4*(4), 131-138.

Hickok, G. (2009). Eight problems for the mirror neuron theory of action understanding in monkeys and humans. *Journal of cognitive neuroscience*, *21*(7), 1229-1243.

Hommel B. Ideomotor action control: On the perceptual grounding of voluntary actions and agents. In: Prinz W, Beisert M, Herwig A, editors, Action Science: Foundations of an Emerging Discipline. Cambridge, MA: MIT Press. 2013. pp. 113–136

Horst, J. S., & Hout, M. C. (2016). The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research. *Behavior Research Methods*, *48*(4), 1393-1409.

Hoshi E, Tanji J. Distinctions between dorsal and ventral premotor areas: anatomical connectivity and functional properties. Curr Opin Neurobiol. 2007;17: 234–242. doi:10.1016/j.conb.2007.02.003

Hugdahl K, Brønnick K, Kyllingsbrk S, Law I, Gade A, Paulson OB. Brain activation during dichotic presentations of consonant-vowel and musical instrument stimuli: a 15O-PET. Neuropsychologia. 1999;37: 431–440. doi:10.1016/S0028-3932(98)00101-8

Hutchinson S, Lee LHL, Gaab N, Schlaug G. Cerebellar volume of musicians. Cereb Cortex. 2003;13: 943–949. doi:10.1093/cercor/13.9.943

Hyde KL, Lerch J, Norton A, Forgeard M, Winner E, Evans AC, et al. The effects of musical training on structural brain development: A longitudinal study. Ann N Y Acad Sci. 2009;1169: 182–186. doi:10.1111/j.1749-6632.2009.04852

Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., & Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science*, *286*(5449), 2526-2528.

Iacoboni, M. (2009). Imitation, empathy, and mirror neurons. *Annual review of psychology*, *60*, 653-670.

Iacoboni M., Molnar-Szakacs I., Gallese V., Buccino G., Mazziotta J. C., Rizzolatti G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biol.* 3:e79. 10.1371/journal.pbio.0030079

Iacoboni M., Woods R. P., Brass M., Bekkering H., Mazziotta J. C., Rizzolatti G. (1999). Cortical mechanisms of human imitation. *Science* 286 2526–2528. 10.1126/science.286.5449.2526

Imada, T., Zhang, Y., Cheour, M., Taulu, S., Ahonen, A., & Kuhl, P. K. (2006). Infant speech perception activates Broca's area: a developmental magnetoencephalography study. *Neuroreport*, *17*(10), 957-962.

Ishida H., Nakajima K., Inase M., Murata A. (2010). Shared mapping of own and others' bodies in visuotactile bimodal area of monkey parietal cortex. *J. Cogn. Neurosci.* 22 83–96. 10.1162/jocn.2009.21185

Jeannerod M. Neural simulation of action: A unifying mechanism for motor cognition. Neuroimage. 2001;14: S103–S109. doi:10.1006/nimg.2001.0832

Julkunen, P., Säisänen, L., Danner, N., Niskanen, E., Hukkanen, T., Mervaala, E., & Könönen, M. (2009). Comparison of navigated and non-navigated transcranial magnetic stimulation for motor cortex mapping, motor threshold and motor evoked potentials. *Neuroimage*, *44*(3), 790-795.

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology*, *3*(3), 203-220.

Kaplan J. T., Iacoboni M. (2007). Multimodal action representation in human left ventral premotor cortex. *Cogn. Proc.* 8 103–113. 10.1007/s10339-007-0165-z

Kiefer, M., Sim, E. J., Herrnberger, B., Grothe, J., & Hoenig, K. (2008). The sound of concepts: four markers for a link between auditory and conceptual brain systems. *Journal of Neuroscience*, *28*(47), 12224-12230.

Kilner J. M. (2011). More than one pathway to action understanding. *Trends Cogn. Sci.* 15 352–357.

Kilner J. M., Friston K. J., Frith C. D. (2007). Predictive coding: an account of the mirror neuron system. *Cogn. Proc.* 8 159–166. 10.1007/s10339-007-0170-2

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35-B42.

Kohler E., Keysers C., Umilta M. A., Fogassi L., Gallese V., Rizzolatti G. (2002). Hearing sounds, understanding actions: action representation in mirror neurons. *Science* 297 846–848.

Kraus N, Chandrasekaran B. Music training for the development of auditory skills. Nat Rev Neurosci. Nature Publishing Group; 2010;11: 599–605. doi:10.1038/nrn2882

Laird AR, Lancaster JL, Fox PT. BrainMap: the social evolution of a human brain mapping database. Neuroinformatics. 2005;3: 65–78. doi:NI:3:1:065

Langheim FJP, Callicott JH, Mattay VS, Duyn JH, Weinberger DR. Cortical systems associated with covert music rehearsal. Neuroimage. 2002; doi:10.1006/nimg.2002.1144

Leaver AM, Lare J Van, Zielinski B, Halpern AR, Josef P. Brain activation during anticipation of sound sequences. J Neurosci. 2010;29: 2477–2485. doi:10.1523/JNEUROSCI.4921-08.2009.Brain

Liberman AM, Mattingly IG. The motor theory of speech perception revised. Cognition. 1985;21: 1–36

Lupyan, G., & Clark, A. (2015). Words and the world: Predictive coding and the language-perception-cognition interface. *Current Directions in Psychological Science*, *24*(4), 279-284.

Macar F, Vidal F, Casini L. The supplementary motor area in motor and sensory timing: Evidence from slow brain potential changes. Exp Brain Res. 1999;125: 271–280. doi:10.1007/s002210050683

Marr D. A Theory of Cerebellar Cortex. J Physiol. 1969;202: 437–470. doi:10.2307/1776957

Martin TA, Keating JG, Goodkin HP, Bastian AJ, Thach WT. Throwing while looking through prisms I. Focal olivocerebellar lesions impair adaptation. Brain. 1996;119: 1183–1198. doi:10.1093/brain/119.4.1183

Matlock, T., Holmes, K. J., Srinivasan, M., & Ramscar, M. (2011). Even abstract motion influences the understanding of time. *Metaphor and Symbol, 26*(4), 260-271.

Mattar, A. A., & Gribble, P. L. (2005). Motor learning by observing. *Neuron*, *46*(1), 153-160.

McDougle, S. D., Ivry, R. B., & Taylor, J. A. (2016). Taking aim at the cognitive side of learning in sensorimotor adaptation tasks. *Trends in cognitive sciences*, *20*(7), 535-544.

McGettigan, C., Walsh, E., Jessop, R., Agnew, Z. K., Sauter, D. A., Warren, J. E., & Scott, S. K. (2013). Individual differences in laughter perception reveal roles for mentalizing and sensorimotor systems in the evaluation of emotional authenticity. *Cerebral cortex*, 227.

Meister, I. G., Krings, T., Foltys, H., Boroojerdi, B., Müller, M., Töpper, R., & Thron, A. (2004). Playing piano in the mind—an fMRI study on music imagery and performance in pianists. *Cognitive Brain Research*, *19*(3), 219-228.

Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., & Iacoboni, M. (2007). The essential role of premotor cortex in speech perception. *Current Biology*, *17*(19), 1692-1696.

Miall R. C. (2003). Connecting mirror neurons and forward models. *Neuroreport* 14 2135–2137. 10.1097/00001756-200312020-00001

Michaels, C. F. (1988). SR compatibility between response position and destination of apparent motion: Evidence of the detection of affordances. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(2), 231.

Mirz F, Ovesen T, Ishizu K, Johannsen P, Madsen S, Gjedde A, et al. Stimulus-dependent central processing of auditory stimuli. A PET study. Scand Audiol. 1999;28: 161–169. doi:10.1080/010503999424734

Mitterschiffthaler MT, Fu CHY, Dalton JA, Andrew CM, Williams SCR. A functional MRI study of happy and sad affective states induced by classical music. Hum Brain Mapp. 2007; doi:10.1002/hbm.20337

Mlynarski W., Hermundstad A. M. (2018). Adaptive coding for dynamic sensory inference. *bioRxiv*[Preprint].

Molnar-Szakacs I., Overy K. (2006). Music and mirror neurons: from motion to‘e’motion. *Soc. Cogn. Affect. Neurosci.* 1 235–241. 10.1093/scan/nsl029

Morrison SJ, Demorest SM, Aylward EH, Cramer SC, Maravilla KR. fMRI investigation of cross-cultural music comprehension. Neuroimage. 2003; doi:10.1016/S1053-8119(03)00300-8

Mueller K, Mildner T, Fritz T, Lepsien J, Schwarzbauer C, Schroeter ML, et al. Investigating brain response to music: A comparison of different fMRI acquisition schemes. Neuroimage. 2011; doi:10.1016/j.neuroimage.2010.08.029

Mukamel, R., Ekstrom, A. D., Kaplan, J., Iacoboni, M., & Fried, I. (2010). Single-neuron responses in humans during execution and observation of actions. *Current biology*, *20*(8), 750-756.

Mushiake H, Inase M, Tanji J. Selective coding of motor sequence in the supplementary motor area of the monkey cerebral cortex. Exp Brain Res. 1990;82: 208–210. doi:10.1007/BF00230853

Nachev P, Kennard C, Husain M. Functional role of the supplementary and pre-supplementary motor areas. Nat Rev Neurosci. 2008;9: 856–869. doi:10.1038/nrn2478

Ohnishi T, Matsuda H, Asada T, Aruga M, Hirakata M, Nishikawa M, et al. Functional anatomy of musical perception in musicians. Cereb Cortex. 2001;11: 754–760. doi:10.1093/cercor/11.8.754

Oliveri, M., Finocchiaro, C., Shapiro, K., Gangitano, M., Caramazza, A., & Pascual-Leone, A. (2004). All talk and no action: a transcranial magnetic stimulation study of motor cortex activation during action word production. Journal of Cognitive Neuroscience, 16(3), 374-381.

Oller, D. K., & Eilers, R. E. (1988). The role of audition in infant babbling. *Child development*, 441-449.

Onuki Y, Van Someren EJ, De Zeeuw CI, Van der Werf YD. Hippocampal-cerebellar interaction during spatio-temporal prediction. Cereb Cortex. 2013;25: 313-321

Öttl, B., Dudschig, C., & Kaup, B. (2017). Forming associations between language and sensorimotor traces during novel word learning. *Language and Cognition*, *9*(1), 156-171.

Patel AD, Iversen JR. The evolutionary neuroscience of musical beat perception: the Action Simulation for Auditory Prediction (ASAP) hypothesis. Front Syst Neurosci. 2014;8: 1–14. doi:10.3389/fnsys.2014.00057

Pellicano, A., Vu, K. P. L., Proctor, R. W., Nicoletti, R., & Umiltà, C. (2008). Effects of stimulus–stimulus short-term memory associations in a Simon-like task. *European Journal of Cognitive Psychology*, *20*(5), 893-912.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., Heisterkamp, S., Van Willigen, B., & Maintainer, R. (2017). Package 'nlme'. *Linear and Nonlinear Mixed Effects Models*, 3-1.

Porro CA, Francescato MP, Cettolo V, Diamond ME, Baraldi P, Zuiani C, et al. Primary motor and sensory cortex activation during motor performance and motor imagery: a functional magnetic resonance imaging study. J Neurosci. 1996;16: 7688–7698. doi:8922425

Prinz W. (2005). A common coding approach to perception and action. In: Neumann O, Prinz W, editors. Relationship between perception and action. Berlin: Springer-Verlag. 1990. pp. 167–201

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature reviews neuroscience*, *6*(7), 576.

Pulvermüller, F., Hauk, O., Nikulin, V. V., & Ilmoniemi, R. J. (2005). Functional links between motor and language systems. *European Journal of Neuroscience*, *21*(3), 793-797.

Raglio A, Galandra C, Sibilla L, Esposito F, Gaeta F, Di Salle F, et al. Effects of active music therapy on the normal brain: fMRI based evidence. Brain Imaging and Behavior. 2015. doi:10.1007/s11682-015-9380-x

Rauschecker JP, Tian B. Mechanisms and streams for processing of "what" and "where" in auditory cortex. Proc Natl Acad Sci. 2000;97: 11800–11806. doi:10.1073/pnas.97.22.11800

Rauschecker JP. An expanded role for the dorsal auditory pathway in sensorimotor control and integration. Hear Res. 2011;271: 16–25. doi:10.1016/j.heares.2010.09.001

Rizzolatti, G., Camarda, R., Fogassi, L., Gentilucci, M., Luppino, G., & Matelli, M. (1988). Functional organization of inferior area 6 in the macaque monkey. *Experimental brain research*, *71*(3), 491-507.

Rizzolatti, G., & Fadiga, L. (1998). Grasping objects and grasping action meanings: the dual role of monkey rostroventral premotor cortex (area F5). *Sensory guidance of movement*, *218*, 81-103.

Rizzolatti, G, Fadiga, L., Gallese, V., Fogassi, L. Premotor cortex and the recognition of motor actions. Curr. Biol. 1996;23: 131–141. doi:10.1016/0926-6410

Rizzolatti G, Fogassi L, Gallese V. Neurophysiological mechanisms underlying the understanding and imitation of action. Nat Rev Neurosci. 2001;2: 661–670. doi:10.1038/35090060

Rizzolatti G, Craighero L. The mirror-neuron system. Annu Rev Neurosci. 2004;27: 169–92. doi:10.1146/annurev.neuro.27.070203.144230

Roest, S. A., Pecher, D., Naeije, L., & Zeelenberg, R. (2016). Alignment effects in beer mugs: Automatic action activation or response competition?. *Attention, Perception, & Psychophysics, 78*(6), 1665-1680.

Rogalsky C, Rong F, Saberi K, Hickok G. Functional anatomy of language and music perception: Temporal and structural factors investigated using functional magnetic resonance imaging. J Neurosci. 2011;31: 3843–3852. doi:10.1523/JNEUROSCI.4515-10.2011

Ross, J. M., Iversen, J. R., & Balasubramaniam, R. (2016). Motor simulation theories of musical beat perception. *Neurocase*, *22*(6), 558-565.

Satoh M, Takeda K, Nagata K, Shimosegawa E, Kuzuhara S. Positron-emission tomography of brain regions activated by recognition of familiar music. Am J Neuroradiol. 2006;27: 1101–1106. doi:27/5/1101

Satoh M, Takeda K, Nagata K, Hatazawa J, Kuzuhara S. Activated brain regions in musicians during an ensemble: A PET study. Cogn Brain Res. 2001; doi:10.1016/S0926-6410(01)00044-1

Satoh M, Takeda K, Nagata K, Hatazawa J, Kuzuhara S. The anterior portion of the bilateral temporal lobes participates in music perception: A positron emission tomography study. Am J Neuroradiol. 2003;24: 1843–1848.

Schmithorst VJ. Separate cortical networks involved in music perception: Preliminary functional MRI evidence for modularity of music processing. Neuroimage. 2005;25: 444–451. doi:10.1016/j.neuroimage.2004.12.006

Schnupp J, Nelken I, King A. Auditory neuroscience. Cambridge, Mass.: MIT Press. 2012

Schubotz, R. I., & Von Cramon, D. Y. (2004). Sequences of abstract nonbiological stimuli share ventral premotor cortex with action observation and imagery. *Journal of Neuroscience*, *24*(24), 5467-5474.

Schubotz RI. Prediction of external events with our motor system : towards a new framework. Trends Cogn Sci. 2007;11: 211–218. doi:10.1016/j.tics.2007.02.006

Scott, S. K., McGettigan, C., & Eisner, F. (2009). A little more conversation, a little less action—candidate roles for the motor cortex in speech perception. *Nature Reviews Neuroscience*, *10*(4), 295-302.

Shepherd S. V., Klein J. T., Deaner R. O., Platt M. L. (2009). Mirroring of attention by neurons in macaque parietal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 106 9489–9494. 10.1073/pnas.0900419106

Simon, J. R. (1969). Reactions toward the source of stimulation. *Journal of Experimental Psychology, 81*(1), 174.

Skipper, J. I., Nusbaum, H. C., & Small, S. L. (2005). Listening to talking faces: motor cortical activation during speech perception. *Neuroimage*, *25*(1), 76-89.

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558-1568.

Snow, J. C., Pettypiece, C. E., McAdam, T. D., McLean, A. D., Stroman, P. W., Goodale, M. A., & Culham, J. C. (2011). Bringing the real world into the fMRI scanner: Repetition effects for pictures versus real objects. *Scientific Reports, 1*, 130.

Snow, J. C., Skiba, R. M., Coleman, T. L., & Berryhill, M. E. (2014). Real-world objects are more memorable than photographs of objects. *Frontiers in Human Neuroscience, 8*, 837.

Spivey, M. J., Tyler, M. J., Richardson, D. C., & Young, E. E. (2000). Eye movements during comprehension of spoken scene descriptions. *Proceedings of the 22nd Annual Conference of the Cognitive Science Society*, 487-492.

Stephan M. A., Lega C., Penhune V. B. (2018). Auditory prediction cues motor preparation in the absence of movements. *Neuroimage* 174 288–296. 10.1016/j.neuroimage.2018.03.044

Stupacher, J., Hove, M. J., Novembre, G., Schütz-Bosbach, S., & Keller, P. E. (2013). Musical groove modulates motor cortex excitability: a TMS investigation. *Brain and cognition*, *82*(2), 127-136.

Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Scifo, P., Fazio, F., Rizzolatti, G., Cappa, S. F., & Perani, D. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *Journal of Cognitive Neuroscience*, *17*(2), 273-281.

Therrien, A. S., Richardson, B. A., & Balasubramaniam, R. (2011). Continuous theta-burst stimulation to primary motor cortex reduces the overproduction of forces following removal of visual feedback. *Neuropsychologia*, *49*(10), 2941-2946.

Therrien, A. S., Lyons, J., & Balasubramaniam, R. (2013). Continuous theta-burst stimulation to primary motor cortex reveals asymmetric compensation for sensory attenuation in bimanual repetitive force production. *Journal of neurophysiology*, *110*(4), 872-882.

Toiviainen P, Alluri V, Brattico E, Wallentin M, Vuust P. Capturing the musical brain with Lasso: Dynamic decoding of musical features from fMRI data. Neuroimage. 2014; doi:10.1016/j.neuroimage.2013.11.017

Trost W, Ethofer T, Zentner M, Vuilleumier P. Mapping aesthetic musical emotions in the brain. Cereb Cortex. 2012;22: 2769–2783. doi:10.1093/cercor/bhr353

Trueswell, J. C., Medina, T. N., Hafri, A., & Gleitman, L. R. (2013). Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive psychology*, *66*(1), 126-156.

Tsai CG, Chen CC, Chou TL, Chen JH. Neural mechanisms involved in the oral representation of percussion music: An fMRI study. Brain Cogn. 2010; doi:10.1016/j.bandc.2010.07.008

Turkeltaub PE, Eden GF, Jones KM, Zeffiro TA. Meta-analysis of the functional neuroanatomy of single-word reading: Method and validation. Neuroimage. 2002;16: 765–780. doi:10.1006/nimg.2002.1131

Tseng Y -w., Diedrichsen J, Krakauer JW, Shadmehr R, Bastian AJ. Sensory prediction errors drive cerebellum-dependent adaptation of reaching. J Neurophysiol. 2007;98: 54–62. doi:10.1152/jn.00266.2007

Umilta, M. A., Kohler, E., Gallese, V., Fogassi, L., Fadiga, L., Keysers, C., & Rizzolatti, G. (2001). I know what you are doing: A neurophysiological study. Neuron, 31(1), 155-165.

Vankov, I. (2011). The interplay of spatial attention and memory: evidence for memory-driven spatial compatibility effects, but not for a facilitating role of attention in memory retrieval. *Attention, Perception, & Psychophysics*, *73*(7), 2124.

Vigneswaran G, Philipp R, Lemon RN, Kraskov A. M1 corticospinal mirror neurons and their role in movement suppression during action observation. Curr Biol. Elsevier Ltd; 2013;23: 236–243. doi:10.1016/j.cub.2012.12.006

Vihman, M. M. (1991). Ontogeny of phonetic gestures: Speech production. In Modularity and the motor theory of speech perception: Proceedings of a conference to honor Alvin M. Liberman (pp. 69-84). Hillsdale, NJ: Erlbaum.

Vukovic, N., Feurra, M., Shpektor, A., Myachykov, A., & Shtyrov, Y. (2017). Primary motor cortex functionally contributes to language comprehension: An online rTMS study. *Neuropsychologia*, *96*, 222-229.

Wallace, V., Menn, L., & Yoshinaga-Itano, C. (1999). Is Babble the Gateway to Speech for All Children? A Longitudinal Study of Children Who Are Deaf or Hard of Hearing. *Volta Review*, *100*(5), 121-48.

Wallmark Z., Deblieck C., Iacoboni M. (2018). Neurophysiological effects of trait empathy in music listening. *Front. Behav. Neurosci.* 12:66. 10.3389/fnbeh.2018.00066

Watkins KE, Strafella AP, Paus T. Seeing and hearing speech excites the motor system involved in speech production. Neuropsychologia. 2003;41: 989–994. doi:10.1016/S0028-3932(02)00316-0

Westermann, G., & Miranda, E. R. (2004). A new model of sensorimotor coupling in the development of speech. *Brain and language*, *89*(2), 393-400.

Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature neuroscience*, *7*(7), 701.

Wilson, M., & Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychological bulletin*, *131*(3), 460.

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science*, *269*(5232), 1880.

Wolpert DM, Kawato M. Multiple paired forward and inverse models for motor control. Neural Networks. 1998;11: 1317–1329. doi:10.1016/S0893-6080(98)00066-5

Wolpert DM, Doya K, Kawato M. A unifying computational framework for motor control and social interaction. Philos Trans R Soc B Biol Sci. 2003;358: 593–602. doi:10.1098/rstb.2002.1238

Woods, E. A., Hernandez, A. E., Wagner, V. E., & Beilock, S. L. (2014). Expert athletes activate somatosensory and motor planning regions of the brain when passively listening to familiar sports sounds. *Brain and cognition*, *87*, 122-133.

Wühr, P., & Ansorge, U. (2007). A Simon effect in memory retrieval: Evidence for the response-discrimination account. *Psychonomic bulletin & review*, *14*(5), 984-988.

Yang, J., & Shu, H. (2016). Involvement of the motor system in comprehension of non-literal action language: a meta-analysis study. *Brain topography*, *29*(1), 94-107.

Yu, C., Smith, L. B., & Pereira, A. F. (2008). Grounding word learning in multimodal sensorimotor interaction. *Proceedings of the 30th Annual Meeting of the Cognitive Science Society*.

Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological science*, *18*(5), 414-420.

Zatorre RJ, Chen JL, Penhune VB. When the brain plays music: auditory–motor interactions in music perception and production. Nature Reviews Neuroscience. 2007;8: 547–558. doi:10.1038/nrn2152

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., & Smith, G. M. (2009). Mixed effects modelling for nested data. In *Mixed Effects Models and Extensions in Ecology with R* (pp. 101-142). Springer, New York, NY.

Zwaan, R. A. (2004). The immersed experiencer: Toward an embodied theory of language comprehension. *Psychology of Learning and Motivation*, *44*, 35-62.

Zwaan, R. A., & Madden, C. J. (2005). Embodied sentence comprehension. *Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thinking*, 224-245.

---------- Forwarded message ---------
From: **Wiley Global Permissions** <permissions@wiley.com>
Date: Fri, May 17, 2019, 3:21 AM
Subject: RE: Permission to include article in dissertation cannot be processed by RightsLink
To: Tim Shea <tim.m.shea@gmail.com>

Dear Tim,

Thank you for your email.

Permission is granted for you to use the material requested for your thesis/dissertation subject to the usual acknowledgements (author, title of material, title of book/journal, ourselves as publisher) and on the understanding that you will reapply for permission if you wish to distribute or publish your thesis/dissertation commercially.

You should also duplicate the copyright notice that appears in the Wiley publication in your use of the Material. Permission is granted solely for use in conjunction with the thesis, and the material may not be posted online separately.

Any third-party material is expressly excluded from this permission. If any material appears within the article with credit to another source, authorisation from that source must be obtained.

Should you require any further information, please do not hesitate to contact me.

Kind regards,

Paisley Chesters
Permissions Co-Ordinator

Wiley
The Atrium
Southern Gate
Chichester
West Sussex
PO19 8SQ
www.wiley.com